J. Korea Saf. Manag. Sci. Vol. 21 No. 3 September 2019
http://dx.doi.org/10.12812/ksms.2019.21.3.011
ISSN 1229-6783(Print)
ISSN 2288-1484(Online)

텍스트 마이닝 기법을 통한 제품 인자 검증 및 안전 관리 연구

정 철 규*·이 창 호*
*인하대학교 산업경영공학과

A Study on the Product Factor Verification and Process Management and Safety Using the Text mining

Chule-kyou Jung*·Chang-Ho Lee*
*Department of Industrial Engineering, INHA University

Abstract

The latest issue is the smart factory. In order to implement this smart factory, the most fundamental element is to establish product specifications for factors affecting the product, obtain useful data to analyzed and predicted, and maintain safety. But most manufacturers have many errors. Therefore, the purpose of this study is to verify factors of product through statistical techniques and to study the process control and safety.

Keywords : SPC, K-means, Word cloud, Text mining, Safety

1. 서 론

우리나라의 제조업은 1970년대 이후 정부의 주도, 수출 중심, 대기업 중심 산업 육성 등으로 국가 경제에 미치는 영향이 세계 최고 수준으로 성장하였다. '12년에는 우리나라 국가 경쟁력에서 제조업이 차지하는 비중은 세계에서 가장 높은 수준으로, 제조업 강국인 독일보다도 8.5%, 일본보다는 10.1%가 높을 정도이다. 그뿐만 아니라 주요 경제지표인 산출액은 물론, 고용에서 제조업이 차지하는 비중 등도 OECD 국가 중 최상위권이다.[1] 하지만 여러 가지 글로벌 경제 상황 및 환경 속에서 제조업의 경쟁력은 낮아지고 있다. 제조업의 경쟁력을 높이기 위한 방법으로 최근 가장 주목되는 이슈는 스마트공장이다. 스마트공장이란 연결성, 가시성, 자기 제어이다. 공장 내 설비와 연계 시스템이 사물 인터넷을 통해 유기적으로 연결된 환경에서 데이터가 실시간으로 수집, 분석되어 공장 내 모든 상황이 일목요연하게 보이고 스스로 제어함으로써 진체 프로세스를 최적화하는 공장으로 볼 수 있다.[2]

제조업에서 이러한 스마트공장을 구현하기 위해서는 가장 중요하고 기본적인 요소는 공정의 안전 및 제품에 영향을 주는 인자에 대한 제품 규격이 설정되고 이 규격을 만족하는 제품을 양산하기 위한 제조 관리기준이 수립되어야 분석이 가능한 유용한 데이터가 축적이 되고 분석 및 예측이 가능해야 한다.

하지만 대부분의 제조업에서는 제품에 영향을 주는 규격에 대한 설정 및 검증, 제조 관리기준 및 인자에 대한 설정부터 잘못된 것들이 많으며, 이러한 설정으로 인하여 이용할 수 없는 데이터들이 누적되고 분석할 수 없는 상황에 있다.

이에 본 연구에서는 실제 화학제품 제조업에서의 제품 규격에 대한 인자들을 통계적 기법을 통해 검증하고 공정 관리 방안에 대한 연구를 목적으로 두고 있다.

2. 이론적 배경

2.1 통계적 공정관리(SPC)

통계적 공정관리는 실험계획, 샘플링검사와 함께 통계적 품질관리의 주된 방법으로 사용되고 있으며, 미국에서 슈하르트가 개발한 관리도에서부터 시작되었다고 할 수 있다. 그 후 전수조사의 대안으로 샘플링검사를 도입하였다.

품질향상 기법은 제조 품질을 나타내는 데이터를 측정

†Corresponding Author: Chang-Ho Lee, Professor, INHA UNIVERSITY, 100, inha-ro, Nam-gu, Incheon, E-mail: lch5601@inha.ac.kr
Received August 16, 2019; Revision September 16, 2019; Accepted September 18, 2019

하는 데에서 시작된다. 데이터를 측정한 다음 공정 상태에 대한 판단과 공정개선을 위한 분석을 해야 하며 이러한 일련의 과정들이 모두 설명될 수 있다. 이런 관점에서 SPC는 공정관리를 위해 사용되었다.[3]

2.2 군집분석

군집분석(Cluster analysis)은 각 개체에 대해 관측된 여러 개의 변수($x_1, x_2, \ldots, x_p$) 값들로부터 n개의 개체를 유사한 성격을 가지는 몇 개의 군집으로 집단화하고, 형성된 군집들의 특성을 파악하여 군집들 사이의 관계를 분석하는 다변량 분석기법이다. 군집분석에 이용되는 다변량 자료는 별도의 반응변수가 요구되지 않으며, 오로지 개체 간의 유사성(Similarity)에만 기초하여 군집을 형성한다.

2.3 계층적 군집(hierarchical clustering)

계층적 군집은 가장 유사한 개체를 묶어 나가는 과정을 반복하여 원하는 개수의 군집을 형성하는 방법이다. 개체 간의 유사성에 대한 다양한 정의가 가능하며, 군집 간의 연결법에 따라 군집의 결과가 달라질 수 있다. 계층적 군집의 결과를 통해 군집들 간의 구조적 관계를 쉽게 살펴볼 수 있다. 이 구조를 통해서 항목 간의 거리, 군집간의 거리를 알 수 있고 군집 내의 항목 간 유사 정도를 파악함으로써 군집의 견고성을 해석 할 수 있다.

계층적 군집을 수행할 때 거리측정 방법에는 최단연결법, 최장연결법, 중심연결법, 평균연결법, 와드연결법이 있다.

1) 유클리드(Euclidean)거리

$$d(i,j) = \sqrt{\sum_{f=1}^{p}(x_{\mathrm{if}} - x_{jf})^2} \quad (1)$$

2) 맨하탄(Manhattan)

$$d(i,j) = \sum_{f=1}^{p}|x_{\mathrm{if}} - x_{jf}| \quad (2)$$

3) 민코우스키(Minkowski)거리

$$d(i,j) = \left|\sum_{f=1}^{p}(x_{\mathrm{if}} - x_{jf})^m\right|^{\frac{1}{m}} \quad (3)$$

4) 표준화(standardized)거리

$$d(i,j) = \sqrt{(x_i - x_j)^T D^{-1}(x_i - x_j)} \quad (4)$$

$D = Diag(S_{11}, \cdots, S_{PP})$: 표본분산(대각)행렬

5) 마할라노비스(Mahalanobis)거리

$$d(i,j) = \sqrt{(x_i - x_j)^T D^{-1}(x_i - x_j)} \quad (5)$$

$S = (S_{ij})_{p\times p}$: 표본공분산행렬

2.4 K-평균군집

K-평균군집(K-means clustering)은 원하는 군집 수 만큼(k개) 초깃값을 지정하고, 각 개체(데이터)를 가까운 초깃값에 할당하여 군집을 형성한 뒤, 각 군집의 평균을 재계산하여 초깃값을 갱신한다. 갱신된 값에 대해 위에 할당 과정을 반복하여 k개의 최종 군집을 형성한다.

2.5 텍스트 마이닝(Text mining)

텍스트 마이닝은 다양한 포맷의 문서로부터 데이터를 획득해 이를 문서별 단어의 매트릭스를 만들어 추가 분석이나 데이터 마이닝 기법을 적용해 통찰을 얻거나 의사결정을 지원하는 방법이다. 다양한 포맷의 문서로부터 텍스트를 추출해 이를 하나의 레코드로 만들어 단어 구성에 따라 데이터마트를 구성하고 이들 간의 관계를 이용해 감성분석이나 World Cloud를 수행하고, 이 정보를 군집화하거나 분류, 사회 연결망 분석에 활용할 수 있다.

2.6 워드클라우드(Word Cloud)

Word cloud는 단어들을 구름 모양으로 나열하여 시각화하는 package이다. 즉 단어를 구름 모양으로 그래픽 화한 워드 클라우드는 빈도가 높은 핵심 단어일수록 가운데에 크게 표현되며, 어떤 단어들을 자주 사용하여 어떤 의미들이 내포되어 있는지 한눈에 알아볼 수 있는 매우 유용한 분석 패키지라 할 수 있다.[5]

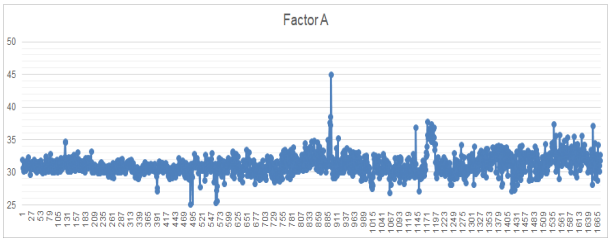
3. 제품 규격 및 인자의 영향

3.1 제품 규격 초과의 원인

화학제품의 규격을 초과하는 현상이 특정 제품군에서 지속해서 발생하였다. 주요 원인으로 화학제품의 투입되는 가공 원재료의 여러 성분 중 A인자의 영향으로 화학제품의 물성 규격이 초과한 것으로 1차원인 조사되었다.

하지만 A인자의 규격은 45 이하로 설정되어 있으며, 제품 규격을 벗어나는 부적합이 발생한 적 또한 없었다. 아래 [Figure 1]같이 원재료의 A인자의 1,644 로트의 분석결과를 시계열로 나열하면 증가하는 경향이 보이며 이 성분의 증가로 인해 화학제품의 물성 부적합이 발생하였다는 R&D 분야의 의견이었다.

ISSN 1229-6783(Print)
ISSN 2288-1484(Online)



[Figure 1] Analysis result of Factor A

3.2 인자의 증가에 대한 관리기준 검토

가공 원재료의 A인자의 증가 원인에 대한 확인을 위해 제품 양산 중 이상 현상 발생 유부에 대한 확인을 진행하였다. 가공 원재료 제조공정 중 관리 기준은 11개로 반응시간, 반응온도, 원재료 투입량, 원재료 분석 결과 등 작업조건에 대한 MES(Manufacturing Execution System) 기록 확인 결과 관리기준에 벗어난 것을 한 건 도 없었다.

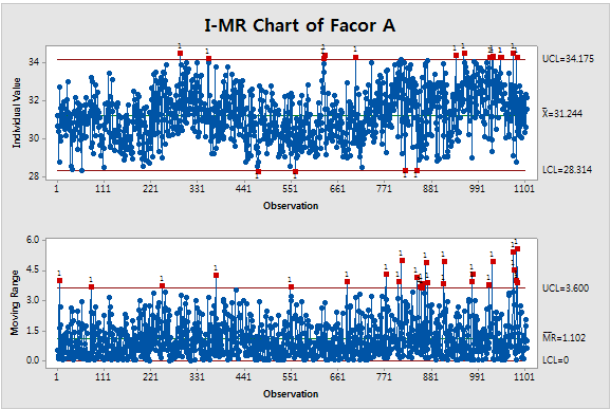
3.3 I-MR 관리도를 통한 모니터링

가공 원재료의 A인자 값의 증가 원인에 대해 공정 중 관리기준에 벗어나거나 이상 현상은 없었다. 하지만 증가 원인을 찾기 위해 I-MR 관리도를 활용하여 원인을 분석하였다.

과거 1,644 로트의 A인자 분석 결과 값을 Minitab 17 version을 이용하여 [Figure 2]의 관리상한(UCL)=34.6, 관리하한(LCL)=28.1을 산출하였다. 이 관리기준을 기준으로 양산 중 이상현상 발생 여부에 대해 모니터링 하였다.

모니터링 중 특정 배치에서 관리상한(UCL)이 벗어난 것이 확인되었으나, 공정 중 관리기준 및 투입 원재료의 성분 및 불순물 값에서 특이사항을 찾아볼 수 없었다.

작업자 인터뷰 시 작업 중에 관리기준을 벗어난 작업은 없었으나 소량의 미용해가 발생하였다는 것을 확인하였다.



[Figure 2] I-MR Chart of factor A

3.4 인자의 대한 군집분석

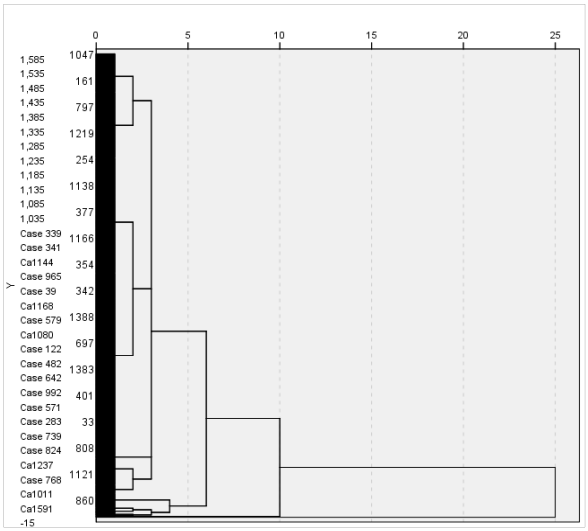
가공 원재료의 A인자에 대한 군집분석을 하여 그룹 간의 특징이 있는지 확인을 위해 분석을 진행하였다. 분석 프로그램은 IBM SPSS Statics 17 Version & MINITAB 17 Version을 사용하였다. A 성분에 대해서 최초 통계프로그램 군집이 나누어지는지 확인하기 위해 초기 군집 분석을 진행하였다. 아래 <Table 1>의 초기 군집분석 결과 중심값 44.97, 24.93의 군집 <Table 2>의 최종 군집분석 시 중심값이 32.88, 30.45의 2개의 군집의 결과를 얻었다. 덴드로그램 확인 시 [Figure 3]과 같이 2개의 군집으로 나타났다.

<Table 1> Initial Cluster Analysis of Factor A

	Cluster	
	1	2
Factor A	44.97	24.93

<Table 2> Final Cluster Analysis of Factor A

	Cluster	
	1	2
Factor A	32.88	30.45



[Figure 3] Dendrogram Analysis of Factor A

초기, 최종 군집 및 [Figure 3]의 덴드로그램 분석을 결과 확인 후 관련 부서 검토를 통해 5개의 군집으로 나누는 것이 유의한 결과를 얻을 것으로 의견이 합의되어 5개의 군집으로 k-means분석을 실시하였다. 5개의 군집에 대해 분석 시 <Table 3>과 같이 초기 군집 분석 결과, <Table 4>, <Table 5>의 최종 군집 분석 결과를 5개의 중심 값이 확인되었다. 분석 결과를 통해 나누어진 군집의

차이점이 있는지 확인하기로 하였다.

<Table 3> Initial Cluster Analysis of Factor A

	Cluster				
	1	2	3	4	5
Factor A	34.3	44.97	39.02	29.60	24.93

<Table 4> Final Cluster Analysis of Factor A

	Cluster				
	1	2	3	4	5
Factor A	32.36	44.97	35.18	30.78	29.25

<Table 5> Final Cluster Analysis Result of Factor A

Cluster	1	418.000
	2	1.000
	3	67.000
	4	840.000
	5	274.000
Valid		1600.000
Missing Value		0.000

3.5 A인자의 텍스트 마이닝 분석

I-MR관리도를 통해 모니터링 중 이상 현상 발생에 대한 작업자 인터뷰 시 소량의 미용해 현상에 대한 의견이 있었다. 작업자 인터뷰 결과를 좀 더 확인하기 위해 원재료 중 A인자의 값이 가장 높은 군집의 원재료 로트를 Word cloud를 통해 인자에 대한 분석을 진행해 보았다.

해당 원재료의 로트는 특정 규칙을 가지고 있지 않았으나, 앞자리는 해당 업체 뒷자리는 포장 단위에 따라 시퀀스로 구분하였다. 로트의 분석을 위해 로트번호의 뒷자리 시퀀스를 제거하는 전처리 작업 후 Word cloud 분석을 하였다.

분석 결과 [Figure 4]와 같이 PD2914, FEB09 등의 특정 로트가 중복적으로 나타나는 것을 확인할 수 있었다.

[Figure 4] Word cloud Analysis of Raw Material LOT

Word cloud 분석을 통해 빈도수가 많은 PD2913, FEB09 로트에 대해 입고 기록을 확인 결과 특정 회사의 제품으로 확인이 되었다. 상기 로트 제품의 경우 [Figure 5] 정상제품, [Figure 6] 미용해 발생 제품같이 입자 크기가 상대적으로 타제품보다 크며 작업 시 일부 미용해가 발생한 것을 확인할 수 있었다.

[Figure 5] Normal working raw material

[Figure 6] Undissolving raw material

미용해 발생 시 로트의 수율 관리를 위해 작업표준에 따라 추가 용해 작업을 위한 약품을 투입하여 작업을 마무리하였으며, A인자의 분석 값 확인 시 규격은 만족하였으나 다른 제품에 비해 높은 수준이 확인되었다. 위의 결과 확인 후 공급 원재료의 특성상 업체 변경 및 제품 규격에 대한 변경은 어려운 상황으로 A인자에 영향을 주는 미용해 발생 시 추가 약품 투입에 대한 작업자 안전을 및 품질을 고려하여 표준작업을 수정하였으며, 개정 후 I-MR관리도를 통해 지속해서 모니터링 시 안정화된 값이 확인되며 이상 현상은 발생하지 않았다.

3.6 제품 규격 초과 인자 검토

화학제품 규격 초과의 원인인 A인자에 대해서 실제 제품의 영향 여부에 대해서 검증을 진행하였다.

k-means 분석을 통해 나누어진 A인자의 5개의 군집을 통해 가장 높은 군과 가장 낮은 군의 결과를 이용하여 동일 화학제품의 물성에 영향을 주는지 Two-Sample T

ISSN 1229-6783(Print)
ISSN 2288-1484(Online)

분석을 하였다. 분석 결과 <Table 6>과 같은 결과를 얻게 되었다. 즉 A인자의 영향으로 화학제품의 물성의 규격 초과 발생의 영향이 없는 것으로 분석이 된다.

<Table 6> Two-Sample T results of Factor A

Two-samp T				
	N	Mean	St Dev	SE Mean
top group	58	5.457	0.280	0.037
sub group	54	5.398	0.209	0.028
Difference = μ(top group) − μ(sub group)				
Estimate for difference : 0.0587				
95% CI for difference : (−0.033, 0.1508)				
T-Test of difference = 0(vs ≠) : T-valude=1.27				
P-value = 0.209				
DF = 105				

3.7 인자에 대한 추가 검증

A인자의 영향에 대해 좀 더 명확하게 확인하기 위해 2가지 제품을 추가로 검증해 보았다. A인자와 화학제품 물성의 결과를 상관 및 회기 분석을 진행하였다.

A화학제품의 분석 결과는 <Table 7>, <Table 8>의 결과를 얻었다. 즉 A인자에 따른 물성 영향은 약한 양의 상관을 보이나 유의하지 않으며, 회귀분석 결과도 설명력이 없으며 유의하지 않았다.

<Table 7> Analysis of Correlation between Product A and Factor A

Pearson correlation = 0.037
P-Value = 0.148

<Table 8> Analysis of Regression between Product A and Factor A

The regression equation					
Y = 6.246 − 0.02603 factor A					
S = 0.266270					
R-Sq = 2.9%					
R-Sq(adj) = 2.0%					
Analysis of Variance					
Source	DF	SS	MS	*F*	*P*
Regression	1	0.19717	0.19717	3.25	0.074
Error	110	6.67140	0.06065		
Total	111	6.86857			

B화학제품의 분석 결과 <Table 9>, <Table 10>의 결과를 얻었다. 이 결과 또한 성분 A에 물성 영향은 약한 음의 상관을 보이나 유의하지 않으며, 회귀분석 결과도 설명력이 없으며 유의하지 않았다.

<Table 9> Analysis of Correlation between Product B and Factor A

Pearson correlation = −0.169
P-Value = 0.074

<Table 10> Analysis of Regression between Product B and Factor A

The regression equation					
Y = 3.555 + 0.03321 Factor A					
S = 1.32959					
R-Sq = 0.1%					
R-Sq(adj) = 0.1%					
Analysis of Variance					
Source	DF	SS	MS	*F*	*P*
Regression	1	3.71	3.71219	2.10	0.148
Error	1561	2759.56	1.76782		
Total	1562	2763.27			

3.8 제품 규격 및 인자의 검토 결과

위의 결과 검토 시 A인자의 영향으로 제품의 부적합이 발생하였다고 결론을 내기 어렵다. 또한 화학제품의 부적합에 따른 최종 제품의 특채를 통해 사용되었을 때, 화학제품의 부적합 영향으로 발생될 고장 모드에 대한 부적합은 발생하지 않았다.

4. 결론 및 향후 연구 과제

본 연구를 통해 실제 화학 제조업에서의 제품 규격에 영향을 주는 인자에 대한 관리기준을 연구를 통해 확인이 보았다. 실제 제품에 영향을 주는 인자에 대한 제품 규격 설정 및 제품 양산 시 규격을 만족하는 제품을 만들기 위해 관리기준 및 안전에 대한 기준 설정 시 실제 접합하지 않은 것들을 확인할 수 있었다. 제품의 설계 단계에서 규격 설정 시 여러 가지 인자 검증 및 유의 수준에 대한 실험계획 같은 방법 등을 통해 제품 규격이 설정되어야 하며, 이러한 규격을 준수하기 위한 관리기준의 선정 및 기준 또한 여러 가지 산업공학적 방법 중 적합한 도구를 이용하여 설정되어야 한다. 또한 불필요한 규격 및 관리기준은 기업

에 많은 비용적인 면에서 낭비가 된다. 또한 화학제품 같은 위험한 요소들이 많은 제조공정 중의 이러한 불합리한 공정관리는 안전에도 큰 영향을 미친다.

4차 산업 시대에 스마트공장 구축을 통한 제조기업의 경쟁력을 갖추기 위해 가장 기본이 되고 중요한 것이 규격에 만족하는 제품을 정확한 관리기준에서 유연하게 만들지고 관리돼야 한다. 이러한 정상적인 상태에서 데이터를 축적하여 예측하고 분석할 수 있어야 한다.

또한 본 연구에서처럼 군집분석 및 텍스트 마이닝 같은 기법을 통해서도 제조업에서 문제 해결 및 최적화에 적용이 가능할 것으로 판단된다. 제품 성적 및 제조 데이터를 군집분석 등을 통해 여러 필요한 공통점 차이점을 확인 수 있으며, 비정형 데이터 및 문자, 로트번호와 같은 기타 공정에 사용되는 데이터를 텍스트 마이닝 등을 통해 공통된 점을 찾는 방법 등을 통해 문제점 등을 해결하는데 유용한 방안이 될 것이며, 이러한 새로운 기법 또한 제조 기업에서 유용하게 활용할 수 있을 것이다.

5. References

[1] Eun Young Kim, Mun Su Park(2018), "A Study on The Limits of Manufacturing Innovation and Policy Direction of Smes in the 4Th Industrial Revolution: Focusing on the Limitations and Examples of Pohang Smes Smart Factory Introduction." Journal of Science & Technology Studies, 18(2):269-306.

[2] Seoung Bum Kim, Seong Hyeon Kang(2016), "The Fourth Industrial Revolution Leading Data Science." Ie Magazine, 9-13.

[3] Chang Soon Park, Jae Heon Lee(2014), "Statistical Process Control." Freeacademy, p.2.

[4] Jong Hwa Nam(2017), "Applied Multivariate Analysis." Freeacademy, 141-151.

[5] Kyoo Sung Noh, Jin Hwa Kim, Seong Taek Park(2016), "Big Data Analytics for Business." p. 149, 160.

저자 소개

정 철 규

순천향대학교 화학공학과 공학사 취득. 인하대학교 대학원 산업경영공학과 석사 취득. 현재 인하대학교 대학원 산업경영공학과 박사 과정 중.
관심분야 : SCM, 품질, 생산관리 등
주 소 : 인천광역시 남구 용현동 253, 인하대학교 산업공학과

이 창 호

인하대학교 산업공학과 학사 취득. 한국과학기술원 산업공학과 석사, 경영과학과 공학박사 취득. 현재 인하대학교 교수로 재직 중.
관심분야 : 물류, RFID, SCM 등
주 소 : 인천광역시 남구 용현동 253, 인하대학교 산업공학과