

코팅제 배합 공정 분석을 위한 빅 데이터-R 기반의 프레임워크를 이용한 데이터 최적화

정영훈*, 이종란*, 노성여**
 *동명대학교 메카트로닉스공학과
 **동명대학교 항만물류시스템학과
 e-mail : sungnsn@naver.com

A Data Optimization using a Framework based on Bigdata-R for Analysis of Coatings Mixing Process

Yeonghoon Jeong*, Jongran Lee*, SeongYeo Noh**
 *Dept. of Mechatronics Engineering, Tongmyong University
 **Dept. of Port Logistics System, Tongmyoung University

요 약

코팅제는 다양한 산업 분야에서 중요성이 더욱 커지고 있으나 실제 제조업체에서는 여전히 작업자의 경험에 의존하여 배합공정을 수행하는 실정이다. 본 논문에서는 R과 실험계획법을 이용하여 코팅제 배합 공정을 분석하기 위한 프레임워크를 제안한다. 제안된 프레임워크를 통한 분석 결과는 보다 정량적인 작업 기준 데이터를 확보하고 작업 현장에 제공함으로써 코팅제 배합 공정을 개선시킬 수 있다. 특히 정확한 배합 기준이 되는 표준 데이터의 부재로 인한 품질 저하와 원가 손실을 감소시키고, 배합 공정에서 발생한 오차 데이터에 대하여 R과 실험계획법을 이용한 분석을 통하여 표준 보정 관계식을 도출함으로써 차후 발생 가능한 오차에 대한 대응 방안을 제시할 수 있다.

1. 서론

코팅제는 도료의 일종으로 제품에 피막층을 형성하여 제품을 보호하고 다양한 특성을 부여하는 고분자 산업에서 활발하게 연구되는 분야 중의 하나이다[1]. 코팅제는 다양한 산업 분야에서 중요성이 더욱 커지고 있으나 실제 제조업체에서는 여전히 작업자의 경험에 의존하여 배합공정을 수행하는 실정이다. 이에 배합 공정 현장에서는 지능화되고 자동화된 처리 기술의 필요성이 대두되고 있다[2, 3]. 본 논문에서는 R과 실험계획법을 이용하여 코팅제 배합 공정을 분석하기 위한 프레임워크를 제안한다. 제안된 프레임워크를 통한 분석 결과는 보다 정량적인 작업 기준 데이터를 확보하고 작업 현장에 제공함으로써 코팅제 배합 공정을 개선시킬 수 있다. 특히 정확한 배합 기준이 되는 표준 데이터의 부재로 인한 품질 저하와 원가 손실을 감소시키고, 배합 공정에서 발생한 오차 데이터에 대하여 R과 실험계획법을 이용한 분석을 통하여 표준 보정 관계식을 도출함으로써 차후 발생 가능한 오차에 대한 대응 방안을 제시할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고 3장에서는 데이터 분석 및 검증 프레임워크 설계를 이용한 코팅제 배합 공정 최적화 TA를 통하여 분석한다. 마지막으로 4장에서는 결론과 향후 연구에 대하여 논의한다.

2. 관련연구

2.1 사물인터넷(IoT, Internet of Things)

사물인터넷이란 사물에 인터넷을 연결하여 사람과 사물이나 사물과 사물간의 정보를 제어할 수 있는 기능이다. 센서 등이 데이터를 수집하고 수집된 데이터를 토대로 임베디드등과 같은 모듈에 데이터를 전송 하게 되면 전송받은 데이터는 네트워크와 통신을 하게 되고 통신을 받은 데이터는 분석하여 처리하게 된다. 처리된 데이터는 응용 서비스 등과 같은 곳에 활용하게 된다. 그림 1은 사물인터넷의 범위 및 요소기술을 보여준다.

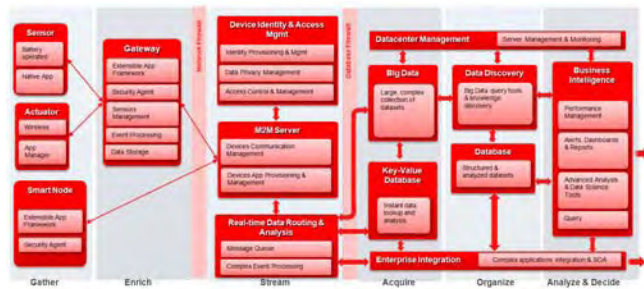


(그림 1) 사물 인터넷의 범위 및 요소기술

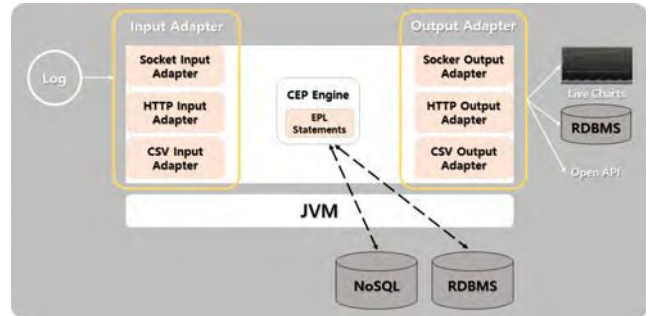
2.2 프레임워크 아키텍처 사례

사물인터넷 시대를 맞이하여 다양한 기관 또는 기업에

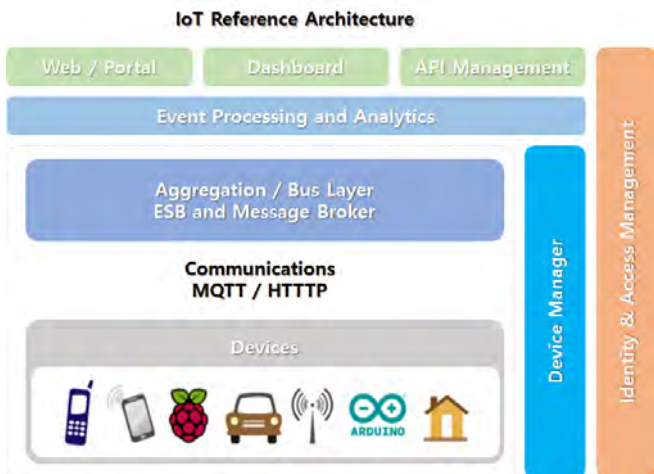
서 IoT를 위한 플랫폼, 프레임워크, 아키텍처들을 제시하고 있다. 이러한 플랫폼을 활용함으로써 좀 더 효율적인 개발이 가능하다. 그림 2는 오라클이 제시하는 IoT 플랫폼의 구성을 보여준다. 그림 3은 오픈소스의 웹서비스 플랫폼을 다루는 기업인 WSO2가 제시한 IoT 레퍼런스 아키텍처이다. 그림 4는 IoT 환경에서 요구되는 기술들을 소개한다[4].



(그림 2) 오라클 IoT 플랫폼 구성[4]



(그림 5) Esper 미들웨어 아키텍처[4]



(그림 3) WSO2의 IoT 레퍼런스 아키텍처[4]

스마트 디바이스	빠른 네트워크	효율적 데이터 프로세싱	대량 데이터 분석	기계 학습
<ul style="list-style-type: none"> · 상황인식 · 스스로 동작 · Power Saving · 네트워크 통신 최소화 	<ul style="list-style-type: none"> · 가능한 많은 (정확한) 데이터를 서버로 전송하기 위한 빠르고 넓은 네트워크 대역 	<ul style="list-style-type: none"> · 데이터에 대한 패턴 감지 · 1차 분석 · Enrichment · 데이터 전달 및 라우팅 	<ul style="list-style-type: none"> · 센서에서 전송한 데이터에 대한 처리 및 분석 · Data Discovery · 비즈니스 정보로 가공 	<ul style="list-style-type: none"> · 분석된 정보를 바탕으로 디바이스 알고리즘 개선 · 스스로 동작하는 스마트 디바이스 발전에 기여

(그림 4) IoT에서 요구하는 기술들[4]

2.3 IoT 미들웨어 기술

IoT 기술 구현을 위해서는 IoT 디바이스와 네트워크 사이의 데이터 처리를 수행하는 IoT 미들웨어가 필요하다. IoT 미들웨어는 네트워크(인터넷)와 사물을 연결해 주는 역할을 수행하며 디바이스, 네트워크, 서비스 등에 대한 추상화 레이어를 제공함으로써 서로 다른 시스템을 연결해 주는 역할을 한다. 복잡한 디바이스의 제어 및 통신 등

여러 기능을 구현한 API를 제공하며 사물인터넷 통신상황 인지 기능 등을 보유하고 있다. 그림 5는 IoT 미들웨어의 하나인 Esper의 아키텍처이다. Esper는 센서에서 나오는 대량 데이터 중에서 정해진 이벤트 패턴을 감지하여 1차가공 후 요약된 결과를 생성(또는 정해진 Action 수행)하고 완성도 높은 오픈소스 프로젝트이다. 상용제품(Oracle Event Processing)의 베이스역할을 한다.

2.4 데이터 처리

BigData는 Hadoop을 중심으로 공개SW 컴포넌트들이 많이 있다. 주로 센서에서 나오는 대용량 비정형 데이터 처리에 적합하며, DW/BI 등과 같은 전통적인 데이터 분석(Analytics) 시스템과 상호 보완적 역할이 가능하다. 그림 6은 빅데이터 분석과 관련된 오픈소스 소프트웨어들을 보여준다.



(그림 6) 빅데이터 분석 관련 오픈소스 소프트웨어[4]

2.5 데이터 분석

데이터분석이란 어떤 기존 시스템 또는 계획 중인 시스템에서 데이터와 데이터의 흐름을 체계적으로 조사하는 것을 의미한다.

데이터통합(Data Integration)은 여러 개의 각기 다른 파일을 결합하면서 중복되는 부분을 완전히 또는 부분적으로 제거하는 것을 의미하며, 데이터통합을 위한 솔루션에는 PentahoData Integration, TalendOpen Studio등과 같은 ETL(Extract Transformation and Load) 솔루션들이

존재한다.

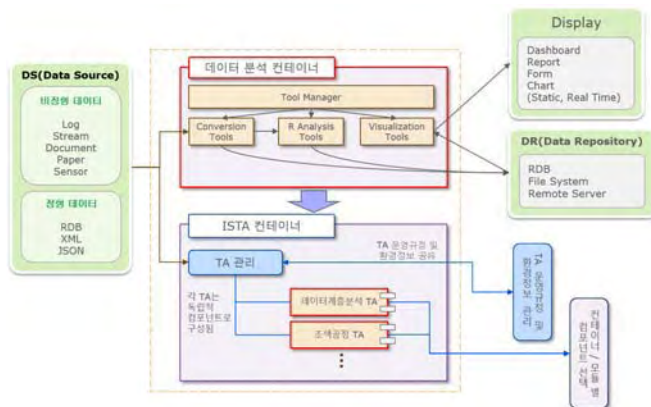
데이터마이닝, 마트를 위한 솔루션에는 PostgreSQL(마트용도), Mondrian(OLAP), Pentaho Data Mining(마이닝) 등의 오픈소스 솔루션들이 존재하며, 특히 웹에서 엄청나게 빠른 속도로 생성되는 웹 페이지(web page) 콘텐츠와 웹 로그(web log), 소셜네트워크서비스의 텍스트 정보와 영상과 같은 비정형 데이터(unstructured data)를 분석하기 위한 다양한 방법론이 등장해 데이터 마이닝의 포괄 범위는 확장되어 가고 있다.

데이터시각화는 QlikView, SpangoBI, PentahoBI 등의 다양하고 완성도 높은 오픈소스 솔루션이 있다. 복잡한 요구사항 수용을 위해서는 d3등과 같은 시각화 라이브러리를 사용하여 개발이 가능하다. 자료로부터 정보를 습득하는 시간 절감으로 즉각적인 상황판단이 가능하며, 자료를 습득하는 사람들의 흥미를 유발하고 정보의 빠른 확산을 촉진한다.

3. 데이터 분석 및 검증 프레임워크 설계를 이용한 코팅제 배합 공정 최적화 TA

3.1 데이터 분석 및 검증 프레임워크

프레임워크의 장점은 소프트웨어 재사용을 통해 비용절감 및 개발기간을 단축할 수 있고, 일관성을 가질 수 있다는 것이다. 본 연구에서는 데이터 분석에 관한 표준 API를 개발하여 사용함으로써 표준 분석 프로세스를 확립하고, 다양한 도메인에서의 요구사항을 반영하였다. 또한 개발하는 프레임워크를 사용하여 각 도메인에 대한 기술구조 확립이 가능하다.



(그림 7) 데이터 분석 프레임워크 구조

개발하는 프레임워크는 데이터 분석을 위한 컨테이너와 산업유형별 표준 구조를 관리하는 ISTA(Industrial Sector Technical Architecture) 컨테이너로 구성된다. 데이터 분석 컨테이너는 데이터 변환 툴(Conversion Tools), R 기반 분석 툴(R Analysis Tools), 시각화 툴(Visualization Tools)로 구성되어 있다.

시스템을 위한 R 기반 분석 모델은 데이터마이닝을 활용한 분석 모델과 작업 데이터의 분석을 통하여 개발하는 신규분석 모델로 나눌 수 있으며, 신규분석 모델 개발은

실험계획법 기반의 분석 모델을 R 통계분석을 활용하여 자체 개발한 데이터 분석 API를 적용하여 데이터를 분석한 후 검증에 필요한 데이터 피드백 과정을 거친 후 설계된다. 데이터마이닝 기법은 기존의 분석 기술과 기계학습 등을 통한 다양한 정량적 분석 및 예측 알고리즘 간의 교차검증과 피드백을 통하여 정성적, 정량적인 예측 결과와 판단근거를 모두 제시할 수 있는 검증 알고리즘을 개발, 적용하여 설계한다.

3.2 코팅제 배합 공정 데이터 분석

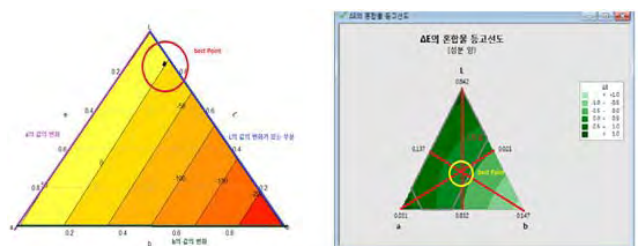
코팅제 제조 시 적용되는 배합 비율 표를 기준으로 정확한 조색을 위하여 보정한 실제 배합 데이터를 이용하여 실험을 수행하였다. 실험데이터는 제조현장에서 직접 측정된 실제 데이터를 적용하고 있으며, 데이터 셋을 제공한 업체의 현장상황에 따라 제조현장의 온도, 습도 등 기초적인 환경데이터가 포함되어 있지 않다. 실제 재료의 배합 시에는 환경적인 요인도 포함되므로 이러한 환경 정보도 함께 수집하여 처리할 필요가 있다. 따라서 실험 결과로도출한 계수와 보정방정식은 환경정보를 적용하여 계속적으로 갱신되어야 한다.

3.2.1 1차 실험 데이터 분석 결과

표 1은 1차 실험에 사용된 코팅제 #01에 대한 표본 데이터의 일부이다. 표 1에서 A4, B4, C1, C2, C4, C8, D3, D7, D14는 코팅제 #01에 투입되는 재료이며, 배합 결과는 CCM을 통한 측정된 후 L, A, B, ΔE로 나타난다. 그림 8은 실험 데이터에 대한 혼합물 등고선도를 MiniTab과 R을 이용하여 계산한 그래프이다.

<표 1> 1차 실험 데이터

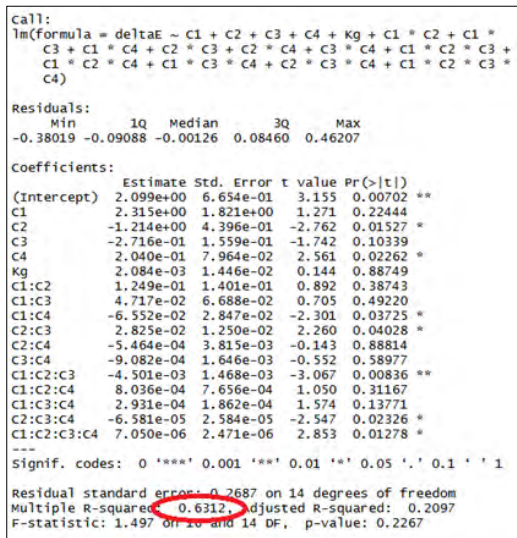
	1	2	3	4	5
C1	42.72	47.32	52.32	57.32	60.72
C2	2.136	2.136	2.136	2.136	2.136
...
C8	58.74	66.34	66.34	73.34	76.34
D3	1.234	2.234	3.934	4.634	4.834
L	-1.51	-0.9	-1.1	-0.52	-0.1
A	3.96	3.06	0.42	-0.22	-0.03
B	-0.16	0.45	-0.82	-0.39	-0.2
E	4.26	3.22	1.43	0.69	0.22



(그림 8) 실험데이터에 대한 혼합물 등고선도

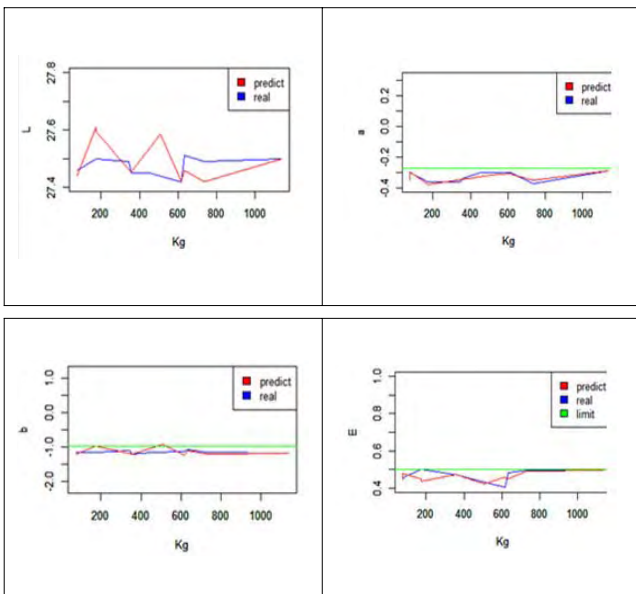
3.3.2 2차 실험 데이터 분석 결과

그림 9는 2차 실험 데이터에 대한 회귀분석을 수행하는 화면이다.



(그림9) R을 이용한 회귀분석 화면

그림 10은 분석결과에 따른 L, A, B, ΔE의 예측 값과 실제 값을 그래프로 보여준다. 파란 선이 실제 데이터의 값이고 빨간 선이 예측된 데이터의 값이다.



(그림 10) L,A,B,ΔE의 예측 값과 실제 값

그림 10은 적합한 색상을 찾아가는 과정이 표시되어 있으며, ΔE가 0.5이하이면 그 색상은 적합한 색상이라고 판단할 수 있으므로 0.5이하의 데이터에 대해서만 비교한 결과, 그림 10과 같은 그래프가 도출되었다. L, A, B에 대한 값을 보여주는 (a), (b), (c) 그래프는 초록색 선에 가까울수록 적합한 색상으로 판단할 수 있고, ΔE 값을 보여주는

(d) 그래프는 초록색 선 보다 밑에 있을 때 적합한 색상으로 판단할 수 있다. 그림 10에서 예측된 결과 값과 실제의 결과 값을 비교해 보았을 때 차이가 근소하며 때로는 예측된 값이 더 좋은 결과를 나타낸다. Lab 그래프는 초록색 선에 가까울수록 적합하고 E 그래프는 초록색 선 보다 밑에 있으면 적합하다.

3.3.3 분석결과

예측된 결과 값과 실제의 결과 값을 비교해보았을 때 차이가 근소하며 때로는 예측된 값이 더 좋은 결과를 나타낸다. 또한 더 많은 데이터를 이용하여 회귀분석을 실시하면 예측에 대한 정확도 또한 증가할 것이다.

4. 결론 및 향후 연구

본 논문에서는 빅 데이터-R 분석을 이용한 실험계획법 기반의 코팅제 배합 데이터 최적화 모형을 제안하였다. 실험을 통하여 실제 생산현장에서 사용하는 코팅제 배합 기준 데이터, 인적 오류, 제품의 특성 값 및 기준 값 등에 발생한 오차에 대하여 작업자가 수정한 보정 결과 데이터를 실험계획법을 기반으로 분석하여 최적의 코팅제 배합 기준 데이터를 계산하였다. 또한 제조 환경 정보와 상황정보를 이용하여 색상과 품질 유지에 가장 중요한 인자를 검색하고 기준 값을 보정하는 최적화 모형을 도출하였다. 본 연구를 통해 확보된 기준 데이터는 제조 공정에 적용할 경우 배합의 정확도 향상과 작업시간 단축을 가능하게 해 주고, 건당 처리시간의 감소로 인한 생산 납품시간 단축, 불량률 감소 등에 따른 원가 절감에 기여할 수 있다. 뿐만 아니라, 다양한 모델링에 대한 제조 공정에서의 표준 데이터를 획득할 수 있다.

향후에는 제안된 최적화 모형의 데이터 처리 아키텍처를 기반으로 개발하는 시스템에 축적된 데이터를 적용하고, 수집된 자료를 데이터 검증 분석 방법에 적용하여 표준 아키텍처의 제시 및 객관적 검증을 수행할 수 있는 기술에 대한 연구에 중점을 두고자 한다.

※본 논문은 중소기업청에서 지원하는 2014년도 산학협력 기술개발사업(No. C0219239)의 연구수행으로 인한 결과물임을 밝힙니다.

참고문헌

- [1] Functional Inorganic Coatings(2007). http://mirian.kisti.re.kr/publication/view.jsp?record_no=155&cont_cd=KI (accessed Nov., 5, 1024)
- [2] Analyst G. Johnson interviewed by S. Bushell in *Computerworld*, on 24 July 2000 ("M-commerce key to ubiquitous internet").
- [3] K. Ashton, "That Internet of Things Thing, in the real world things matter more than ideas," *RFID Journal*, 2009.
- [4] 정명훈, "사물인터넷(Internet of Things) 시대의 공개 SW," *제64회 Open Technet*, 2014.