

## Optimization Model for the Mixing Ratio of Coatings Based on the Design of Experiments Using Big Data Analysis

Noh Seong Yeo<sup>†</sup> · Kim Young-Jin<sup>††</sup>

### ABSTRACT

The research for coatings is one of the most popular and active research in the polymer industry. For the coatings, electronics industry, medical and optical fields are growing more important. In particular, the trend is the increasing of the technical requirements for the performance and accuracy of the coatings by the development of automotive and electronic parts. In addition, the industry has a need of more intelligent and automated system in the industry is increasing by introduction of the IoT and big data analysis based on the environmental information and the context information. In this paper, we propose an optimization model for the design of experiments based coating formulation data objects using the Internet technologies and big data analytics. In this paper, the coating formulation was calculated based on the best data analysis is based on the experimental design, modify the operator with respect to the error caused based on the coating formulation used in the actual production site data and the corrected result data. Further optimization model to correct the reference value by leveraging big data analysis and Internet of things technology only existing coating formulation is applied as the reference data using a manufacturing environment and context information retrieval in color and quality, the most important factor in maintaining and was derived. Based on data obtained from an experiment and analysis is improving the accuracy of the combination data and making it possible to give a LOT shorter working hours per data. Also the data shortens the production time due to the reduction in the delivery time per treatment and It can contribute to cost reduction or the like defect rate reduced. Further, it is possible to obtain a standard data in the manufacturing process for the various models.

**Keywords :** Big Data, Design of Experiments, Mixing of Coatings, Data Optimization

## 빅데이터 분석을 활용한 실험계획법 기반의 코팅제 배합비율 최적화 모형

노성여<sup>†</sup> · 김영진<sup>††</sup>

### 요 약

코팅제에 대한 연구는 고분자 산업에서 가장 보편화되고 활발하게 연구되고 있는 내용의 하나이다. 코팅제는 전자산업, 의료, 광학 분야 등에서 중요성이 더욱 커지고 있으며, 특히 자동차 및 전자부품의 첨단화에 힘입어 코팅제에 대한 성능과 정밀도 등 기술적인 요구사항이 증가하고 있는 추세이다. 또한 방대한 환경 정보와 상황 정보를 기반으로 한 사물 인터넷과 빅데이터 분석 기술의 도입을 통해 산업 현장에서는 더욱 지능화되고 자동화된 시스템과 처리 기술의 필요성이 높아지고 있다. 이에 본 논문에서는 사물 인터넷 기술과 빅데이터 분석을 활용한 실험계획법 기반의 코팅제 배합 데이터에 대한 최적화 모형을 제안한다. 본 논문에서는 실제 생산현장에서 사용하는 코팅제 배합 기준 데이터와 발생한 오차에 대하여 작업자가 수정한 보정 결과 데이터를 실험계획법을 기반으로 분석하여 최적의 코팅제 배합 기준 데이터를 계산하였다. 또한 빅데이터 분석 기술과 사물 인터넷 기술을 활용하여 기존의 코팅제 배합 기준 데이터만을 적용한 공정이 아니라 제조 환경 정보와 상황 정보를 이용하여 색상과 품질 유지에 가장 중요한 인자를 검색하고 기준값을 보정하는 최적화 모형을 도출하였다. 실험 및 분석을 통해 확보된 기준 데이터는 제조 공정에 적용할 경우 배합의 정확도 향상과 LOT별 작업시간 단축을 가능하게 해주고, 긴당 처리시간의 감소로 인한 생산 납품시간 단축, 불량률 감소 등에 따른 원가 절감에 기여할 수 있다. 또한, 다양한 모델링에 대한 제조 공정에서의 표준 데이터를 획득할 수 있다.

**키워드 :** 빅데이터, 실험계획법, 코팅제 배합, 데이터 최적화

\* 이 논문은 2014년도 산업통상자원부의 지원으로 기술경영 전문인력 양성사업의 지원을 받아 수행된 연구임(1415134318).

† 준회원: 부경대학교 기술경영협동과정 박사과정

†† 비회원: 부경대학교 시스템경영공학부 교수

Manuscript Received: September 5, 2014

Accepted: October 16, 2014

\* Corresponding Author: Kim Young-Jin(youngk@pknu.ac.kr)

### 1. 서 론

도료란 물체의 표면에 피막층을 형성함으로써 물체를 보호하고 다양한 성능을 부여하는 화학제품을 말한다. 다양한

도료 중에서도 특히 코팅제에 대한 연구는 고분자 산업에서 가장 보편화되고 활발하게 연구되고 있는 내용의 하나이다. 그러나 국내에서는 많은 기업들이 코팅제에 대하여 관심과 노력을 집중하고 있음에도 불구하고 이미 국내 시장의 많은 부분이 외국기업에 의해 잠식되고 있는 실정이다. 국내와 달리 선진 외국기업들은 원료의 배합 기술 수준을 넘어 새로운 형태의 제품을 개발하기 위한 시도를 하고 있다[1].

최근의 기술 연구는 다양한 산업의 융합이 주류를 이루고 있다. 코팅제의 경우는 전자산업, 의료, 광학 분야 등에서 중요성이 더욱 커지고 있으며, 특히 자동차 및 전자부품의 첨단화에 힘입어 코팅제에 대한 성능과 정밀도 등 기술적인 요구사항이 증가하고 있는 추세이다.

또한 방대한 환경 정보와 상황 정보를 기반으로 한 사물 인터넷(IoT, Internet of Things)[2, 3]과 빅데이터[4, 5] 분석 기술의 도입을 통해 산업현장에서는 더욱 지능화되고 자동화된 시스템과 처리 기술의 필요성이 높아지고 있다. 이에 본 논문에서는 빅데이터 분석을 활용한 실험계획법 기반의 코팅제 배합 데이터에 대한 최적화 모형을 제안한다.

코팅제는 다양한 산업 분야에서 활용되는 중요한 재료의 하나이며 완성된 제품의 성능과 디자인적인 요소의 향상을 위하여 정확한 색상과 품질이 유지되어야 한다. 그러나 실제 현장에서는 여전히 경험을 기반으로 한 수작업에 의해 배합이 이루어지고 있는 실정이다. 또한 정확한 표준 배합 데이터의 미확립으로 품질의 저하 및 제조 과정에서 원재료 손실을 유발하고 있으며, 생산 단위의 크기에 따라 기준에 제공되는 기준 배합 비율표와 생산 결과가 일치하지 않는 경우도 많이 발생하고 있다.

본 논문에서는 실제 코팅제 배합 기준 데이터와 발생한 오차에 대하여 작업자가 수정한 보정 결과 데이터를 실험계획법을 기반으로 분석하여 최적의 코팅제 배합 기준 데이터를 확보한다. 또한 빅데이터 분석 기술과 사물 인터넷 기술을 활용하여 기존의 코팅제 배합 기준 데이터만을 적용한 공정이 아니라 제조 환경 정보와 상황 정보를 이용하여 색상과 품질 유지에 가장 중요한 인자를 검색하고 기준값을 보정하는 최적화 모형을 도출하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 제안하는 최적화 모형과 관련된 연구들을 소개한다. 3장에서는 코팅제 배합 최적화 모형의 설계를 다루고, 4장에서는 제안 모형의 검증과 분석 내용을 실험을 통해서 설명한다. 마지막으로 5장에서는 결론과 향후 연구에 대하여 논의한다.

## 2. 관련 연구

### 2.1 빅데이터 분석

빅데이터[4, 5]는 부피가 크고(Volume), 변화의 속도가 빠르며(Velocity), 데이터의 속성이 다양한 데이터(Variety)를 지칭한다. 빅데이터는 비정형 텍스트 데이터에서 정보를 추출, 가공하는 텍스트 마이닝, 소셜 미디어 텍스트의 긍정, 부정, 중립의 선호도 판별에 사용되는 오피니언 마이닝, 소셜

네트워크의 연결 구조 및 강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하여 활용되는 소셜 네트워크 분석, 비슷한 특성을 가진 개체를 합쳐가면서 유사 특성의 군집을 발굴하는 군집 분석 등의 분석 인프라를 가진다[6].

기존의 빅데이터 분석 시스템에서 많이 활용되고 있는 하둡 프레임워크를 기반으로, 제안하는 시스템을 위한 모듈의 개선 및 확장, 신규 개발을 수행한다. 개발하는 모듈은 확장 HDFS 저장소, 비정형 데이터 변환 모듈 등을 포함한다. 데이터의 종류에는 정형 데이터와 비정형 데이터가 있는데 특히, 비정형 데이터를 중심으로 데이터를 수집, 저장, 관리하며 분석에 활용한다.

### 1) HDFS(Hadoop File System)의 구조[7]

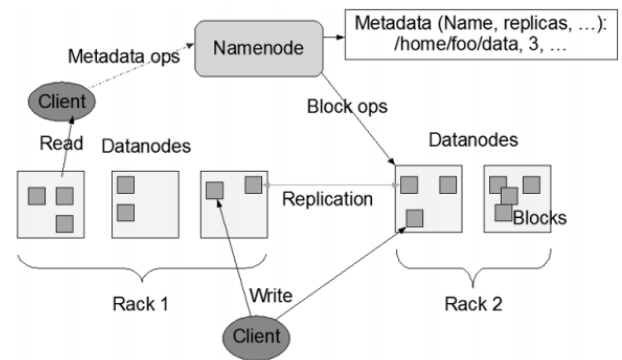


Fig. 1. The Architecture of HDFS

- 블록 : 기본 단위, HDFS파일들은 블록 크기의 청크로 쪼개지고 독립적인 단위로 저장된다.
- 탐색 비용을 최소화하기 위하여 블록의 크기가 커질 수 있다.
- 분산 파일 시스템은 블록에 대한 추상 개념을 통해 몇 가지 이점을 얻을 수 있다.
  - a. 파일 하나가 네트워크에 있는 어떤 하나의 디스크보다도 더 커질 수 있다. 하나의 파일을 구성하는 블록이 동일 디스크에 저장될 필요가 없기 때문에 블록은 클러스터 내의 어떠한 디스크에도 저장될 수 있다.
  - b. 추상화 단위로 파일보다 블록을 사용하는 것이 저장소의 서브시스템을 단순하게 만들며 블록은 복제를 효율적으로 수행할 수 있다.
- 네임 노드와 데이터 노드
  - a. NameNode는 HDFS의 마스터 데이터가 되며 슬레이브 데이터인 DataNode에게 I/O 작업을 지시한다.
  - b. NameNode는 어떻게 파일이 블록 단위로 나누어져 있는지, 어느 노드가 해당 블록을 가지고 있는지와 분산 파일 시스템의 전반적인 상태를 저장한다.
  - c. NameNode는 주로 메모리와 I/O에 관한 일을 담당한다.

- d. NameNode를 담당하는 서버는 작업량을 덜기 위해 데이터 저장 및 MapReduce 프로그램을 위한 어떤 연산 작업도 담당하지 않는다.
- e. 클러스터에 포함되어 있는 슬레이브 머신은 DataNode를 포함하며, 이것은 로컬 파일 시스템에 위치한 파일에 HDFS 블록을 기록하거나 해당 파일을 읽는 등의 단순한 기능을 수행한다.

2) Flume(데이터 수집 모듈)(8)

- 데이터 발생원으로부터 데이터를 수집해 하둡 파일 시스템에 안정적으로 저장한다.
- Flume은 단순하며 유연한 스트리밍 데이터 플로우(Streaming Data Flow) 아키텍처를 기반으로 한다. 또한 장애 발생 시 쉽게 대처할 수 있으며, 로그 유실에 대한 신뢰 수준을 상황에 맞게 변경할 수 있을 뿐만 아니라, 다양한 복구 메커니즘을 제공한다. 실시간 로그 분석 애플리케이션을 개발할 수 있도록 간단하고 확장 가능한 데이터 모델을 사용한다. 클러스터에 있는 모든 장치로부터 로그 파일들을 수집한 후, 하둡 분산 파일 시스템(HDFS)과 같은 중앙 저장소에 저장하는 로깅 시스템을 구축해야 할 때 안정맞춤이다. Flume은 운영가능성(Manageability), 확장성(Scalability), 신뢰성(Reliability), 확장성(Extensibility)을 핵심 목표로 두고 만족시키도록 만들어졌다.

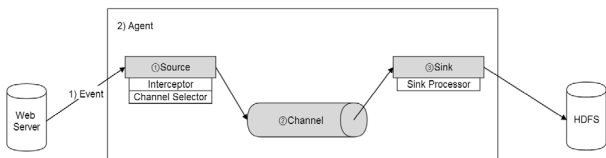


Fig. 2. The Event Flow of Flume

- Flume Agent : 이벤트를 전달하는 컨테이너로서 Source, Channel, Sink로 흐름을 제어하며, Agent 간 이벤트 이동이 가능하다. 1개의 Agent가 다수의 Agent로 연결할 수 있다.
  - a. Source : 이벤트를 수집하여 한 개 이상의 채널로 전달한다. Interceptor, Channel Selector를 통해 수집된 데이터를 변경하거나 Channel을 지정할 수 있는 기능도 제공한다.
  - b. Channel : 이벤트를 Source와 Sink로 전달하는 버퍼이다.
  - c. Sink : Channel로부터 전달받은 이벤트를 저장, 전달한다. Sink Processor를 통해 Sink할 대상을 다중 선택하거나 여러 개의 Sink를 하나의 그룹으로 관리하여 Failover(장애 극복)에 대응할 수 있다.

제조업에서는 하드웨어적 성장의 한계 극복을 위하여, 전수 데이터에 가까운 엄청난 양의 정보를 사용할 수 있게 되는 빅데이터 분석 기술이 각광받고 있다. 빅데이터가 각광

을 받는 것은 센서, 클라우드, 소프트웨어, 하드웨어 기술들이 잘 뒷받침되고 융합될 수 있었기 때문이다. 빅데이터는 성장한계를 경험하고 있는 제조업에 폭넓게 적용되어 생산성 향상 측면뿐 아니라 제조업의 서비스화 및 첨단화 측면에 있어서 산업 전반에 미칠 수 있는 잠재력이 크다고 할 수 있다.

2.2 R을 이용한 빅데이터 분석

R[9]은 통계적 프로그래밍과 시각화에 효과적인 오픈소스 프로젝트이다. 데이터와 관련된 많은 입출력, 핸들링, 관리, 분석, 그래픽 등 최신의 알고리즘과 라이브러리를 제공하며 이미 수백만 명 이상의 사용자를 보유하고, 수천 명의 기여자에 의한 각 업무 도메인과 관련된 리소스 및 도움말을 제공하고 있다. 특히 Google과 Facebook은 R을 자사의 주된 분석 플랫폼으로 활용 중이다. R은 빅데이터 분석을 위한 아키텍처 전반에 걸쳐 공통적인 분석 플랫폼으로 자리 잡고 있으며 빠른 처리속도를 위한 In-Memory 컴퓨팅, 분석과 시각화 작업을 위한 강력한 그래픽 기능 지원과 같은 특징을 가진다.

R을 이용한 데이터 분석은 빅데이터가 이슈화되기 전부터 많이 사용되어 왔지만, 기본적으로 빌드된 R에서는 단 하나의 코어 프로세스만을 사용하며, 전체적으로 느린 속도를 보이고, 또한 32비트로 작성된 코드에 의해 메모리의 한계가 2GB로 지정되어 있는 문제 때문에 빅데이터 분석에는 어렵다는 문제가 있었다. 그러나 R 2.14 버전 이후부터는 바이트 컴파일러의 채용으로 인해 연산속도가 약 5배가량 증가하였으며, 일부 병렬처리가 가능한 함수에 대하여 멀티코어 지원을 허용함으로써 많은 부분이 해결되었다. 메모리 한계의 문제는 64비트용 코드를 적용함으로써 일부는 개선되었지만 64비트에 최적화된 자료형이 부족하여 일부 문제는 여전히 존재한다. 그러나 R만을 사용하지 않고 데이터베이스와 연계하여 분석을 수행할 경우 다소 불편함은 남아 있지만 메모리의 한계에 대한 문제는 해결가능하다[10].

Fig. 3은 개발하는 시스템에 적용될 R 기반의 데이터 분석 모델의 종류와 해당 분석모델을 이용하여 TA를 생성하는 과정을 보여주고 있다. 적용하고자 하는 대표적인 분석모델로는 선형회귀분석, 군집분석, 시계열분석 등을 들 수 있다.

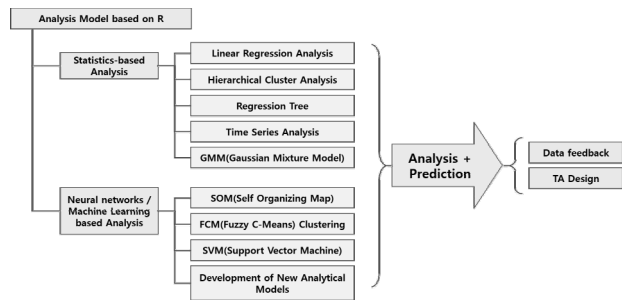


Fig. 3. The Process of Data Analysis Models using R base Platform for Developed System

선형회귀분석은 선형성이라는 기본 가정이 충족된 상태에서 독립변수와 종속변수의 관계를 설명하거나 예측하는 통계방법이다. 독립변수가 하나인 경우를 단순회귀분석, 여러 개인 경우를 다중회귀분석이라고 한다. 즉 회귀분석에서 독립변수에 따라 종속변수의 값이 일정한 패턴으로 변해 가는데, 이러한 변수 간의 관계를 나타내는 회귀선이 직선에 가깝게 나타나는 경우를 선형회귀분석이라고 한다[11].

군집분석이란 데이터 간의 유사도를 정의하고 그 유사도에 가까운 것부터 순서대로 합쳐가는 방법으로, 유사도의 정의에는 거리나 상관계수 등 여러 가지가 있다. 군집분석에는 차례대로 합쳐가는 계층적 방법 이외에, 요인분석 등으로 미리 군집을 예상하여 합쳐가는 비계층적 방법도 있다[12].

시계열 데이터는 어떤 현상에 대하여 시간의 흐름에 따라 일정한 간격으로 관측하여 기록한 데이터를 의미한다. 시계열분석이란 시계열 데이터에 바탕을 둔 분석 방법으로, 시계열 데이터가 시간에 따라 변화하는 구조 사이에 존재하는 다양한 확률적, 통계적 성질을 파악하고, 이를 바탕으로 확률 현상의 미래 형태를 예측하는 통계분석을 의미한다[13].

2.3 조색 원료 배합

코팅제를 포함한 도료의 구성성분은 수지, 안료, 용제, 첨가제로 나눌 수 있다. 수지는 도료의 기본적인 성분으로 도막의 최종 성능인 화학적, 물리적 기능을 좌우하게 된다. 흔히 피착제(substrate)에 도포되어 필름을 형성시키는 물질이며 도료 성분에서 요구되는 물성을 좌우하는 성분이기 때문에 다양한 고분자들이 도료용 수지로 이용되고 있다.

안료는 용제에 용해되지 않고 분산되어 있는 옥사이드나 실리케이트와 같은 무기 화합물을 이르는 총칭이다. 이러한 안료의 경우 프라임(prime) 안료와 체질 안료로 구분이 가능하다. 프라임 안료는 색깔을 부여하거나 불투명하게 만들고, 예폭시와 같이 빛에 민감한 수지를 자외선으로부터 보호하는 기능을 줄 수 있다. 체질안료의 경우 색을 부여하거나 불투명성을 만드는 것과 관계없이 도막의 기계적인 성능을 개선시키고 도료의 가격을 낮추는 긍정적인 작용을 한다. 이외에 안료는 내열성, 방오성과 같은 다양한 형태의 기능성 안료들도 존재한다. 안료는 조색제라고도 표현한다[1].

첨가제는 도료의 제조부터 도료가 건조되어 내구성을 가질 때까지 각각 여러 기능을 발휘하도록 첨가되어지는 보조적 기능을 가진 물질을 총칭한다[1].

단일 안료를 사용한 도료를 원색이라 하고 이러한 원색들을 서로 혼합하여 원하는 색상으로 조합하는 작업을 조색이라 한다. 조색에는 육안조색법과 CCM(Computer Color Matching)[15] 활용조색이 있다. 육안조색의 경우에는 조색자의 숙련된 기술을 필요로 하며, CCM 활용조색의 경우에는 컴퓨터 매칭 시스템을 사용하게 되므로 과학적이고 체계적인 조색을 하는 데 유용하다. Fig. 4는 기존 코팅제 배합 제조 공정의 흐름도이다.

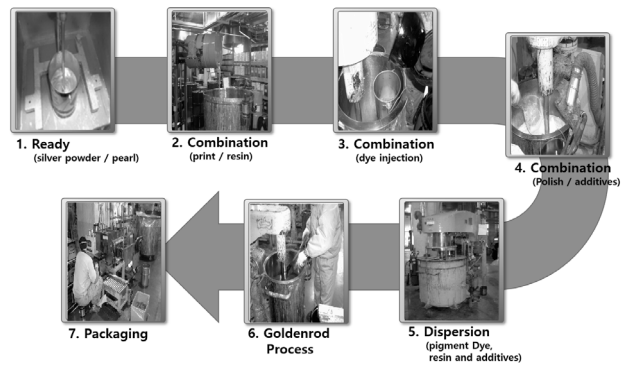


Fig. 4. The Manufacturing Process for Typical Coatings Mixing

2.4 실험 계획법(DOE, Design of Experiments)

실험계획법이란 실험을 어떻게 수행하고, 데이터를 어떻게 수집하며, 데이터를 어떤 통계적 방법으로 어떻게 분석하면 최소의 실험만으로 최대의 정보를 얻을 수 있을지에 대하여 계획하는 방법을 말한다. 실험계획법은 영국의 통계학자 R. A. Fisher가 로담스테드 실험 농장에서 작물 수확에 영향을 주는 주요 인자를 연구하기 위한 방법으로 구체화하였으며, 농업뿐만이 아니라, 기계, 생물, 전자, 화학, 제약, 식품 등 다양한 산업에서 연구개발 및 품질 개선을 위한 도구로서 활용되고 있다[16, 17, 18]. 실험계획법의 설계 시 사용되는 용어들은 Table 1과 같다. 실험계획법에서 실험대상이 되는 프로세스를 그림으로 나타내면 Fig. 5와 같다.

Table 1. The Terms used in the Design of Experiments[19]

Terms	Description
Characteristics	Results, measured in the form of data from the experiment
Factor	The causes that are treated in the experiment directly in a number of causes that is thought to influence the characteristic value
Levels	Terms of factors for the Experimental
Repeat	That at the same level as the combination of factors to be treated in this experiment, the experiment is performed more than once

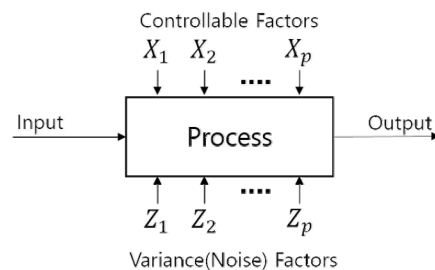


Fig. 5. Model of the Experiment Process

실험계획법은 랜덤화, 반복, 블록화의 세 가지 기본 원리로 구성된다. 랜덤화(Randomization)는 실험 단위를 결정하거나 실험 순서를 임의의 값으로 결정하는 것으로, 선택되

지 않은 요인들이 실험 결과에 편향된 영향을 미치는 것을 막기 위한 방법이다[19].

반복(Replication)은 인자들을 동일한 조건하에서 두 번 이상 실험을 수행하는 것을 뜻한다. 동일한 조건하에서 수행한 실험이라도 제어 불가능한 오차가 항상 수반되기 때문에 실험 결과가 반드시 동일하지는 않다.

블록화(Blocking)는 실험 전체를 뭉 수 있는 한 동질적인 부분으로 나누어 여러 블록으로 구성한 후, 각 블록 내에서 인자효과를 조사하는 것을 말한다.

실험계획법은 대표적으로 요인배치법(Factorial Design), 분할법(Split-Plot Design), 교락법(Confounding Method), 일부실시법(Fractional Factorial Design), 불완비블록계획법(Incomplete Block Design), 반응표면계획법(Response Surface Design), 혼합물 실험계획법(Mixture Design), 로버스트 실험계획법(Robust Design)으로 분류할 수 있다.

이러한 여덟 가지의 분류 중 본 논문에서는 요인배치법과 혼합물 실험계획법을 이용하여 실험을 수행한다.

#### 1) 요인배치법(Factorial Design)[19]

요인배치법은 일원배치법, 이원배치법, 다원배치법으로 구분된다. 이원배치법은 반응값에 영향을 줄 것으로 기대되는 인자의 수가 2개이고 실험단위가 동질적이어서 실험의 랜덤화에 제약조건이 없는 실험계획법을 말한다. 반복이 없는 것과 있는 것으로 구분할 수 있으며 상호작용의 차이를 확인하여야 한다. 이때, A요인과 B 요인이 상호작용을 하지 않는다는 가정을 필요로 한다. 그렇지 않은 경우에는 각 요인이 종속변수에 미치는 영향이 명확하게 나타나지 않는다. 이원계획법의 각 표본들은 독립이어야 하며, 모든 행과 열을 조합에 대해 모집단이 정규분포를 이루고, 분산이 같아야 한다는 조건을 만족하여야 한다.

인자의 수가 3개 이상인 경우는 다원배치법이라고 한다.

#### 2) 혼합물 실험계획법(Mixture Design)[19]

대부분의 실험계획법은 하나 또는 두 개 이상의 인자  $(x_1, x_2, \dots, x_k)$ 가 어떤 관심 있는 반응값  $y$ 에 유의한 영향을 미치는가를 발견하거나, 더 나아가서  $y$ 를 최대 또는 최소화시키는  $x_i$ 들의 최적조건을 찾는 데 그 목적이 있다. 일원배치법, 이원배치법, 분할법, 반응표면 실험계획법 등이 여기에 속한다.

그러나 잉크, 타이어, 페인트 등과 같이 제품이 여러 개의 성분의 혼합으로 이루어져 있고 각 성분의 혼합비율만이 문제가 되는 경우, 어떠한 성분이 관심이 있는 반응값에 유의한 영향을 미치는지, 반응을 최대 또는 최소로 만드는 최적 혼합비율이 무엇인지 찾고자 하는 실험계획을 혼합물 실험계획법이라고 부른다.

$k$ 개 성분의 혼합에서  $x_i$ 를  $i$ 번째 성분의 혼합비율이라고 하면 식 (1)과 같은 관계식을 만족해야 한다.

$$x_1 + x_2 + \dots + x_k = 1, (x_i \geq 0, i = 1, 2, \dots, k) \quad (1)$$

혼합물 실험계획법에서는 혼합비율의 최적조건을 찾기 위해 회귀식의 기법을 이용하며 주로 사용하는 모형은 1차, 2차 회귀모형이다.

#### 2.5 사물 인터넷(IoT, Internet of Things)[2, 3]

사물 인터넷(IoT)이란 인터넷을 기반으로 모든 사물을 연결함으로써 사람과 사물, 사물과 사물 간의 상호 정보 교환을 통해 구축할 수 있는 다양한 지능형 기술과 서비스를 말한다. 1999년 메사추세츠 공과대학(MIT)의 Auto-ID Center의 Kevin Ashton에 의하여 미래에는 RFID와 다양한 센서를 일상생활에 사용하는 사물에 탑재하여 서비스를 제공하는 사물 인터넷 시장이 열릴 것으로 전망하면서 처음 사용되었다.

IoT는 인터넷에 연결된 기기가 사람의 개입 없이 상호간에 정보를 교환하고 처리한다. 기존의 유비쿼터스 환경(Ubiquitous Environment)이나 사물통신(M2M, Machine to Machine)과 유사하다고 볼 수 있으나 M2M은 장비간의 통신을 통해 가공된 정보를 사람에게 제공하는 수동적인 개념을 인터넷으로 확장하여 모든 정보와의 상호작용을 수행하는 개념으로 발전한 M2M의 후속 단계로 볼 수 있다.

IoT를 구현하기 위한 기반 기술로는 주위의 장비와 환경으로부터 정보를 수집하는 센싱기술, 사물을 인터넷에 연결해주는 유무선 통신/네트워크 인프라 기술, 다양한 서비스와 형태에 적합하게 정보를 가공, 처리하고 기술을 융합하도록 해주는 서비스 인터페이스 기술을 들 수 있다.

### 3. 코팅제 배합 데이터 최적화 모형의 설계

#### 3.1 데이터 셋의 구성 및 분석

본 논문에서는 코팅제 제조 시 적용되는 기준 배합 비율표를 기준으로 정확한 조색을 위하여 보정한 실제 배합 데이터를 실험에 적용하였다. 조색이란 문자 그대로 여러 가지 채색을 조합하여 만들고자 하는 빛깔을 나타내는 것이다. 일반적으로 코팅제의 배합에는 기준 배합 비율표가 제시되지만, 실제 배합 시에는 배합하려는 총량과 배합 시의 습도, 온도와 같은 환경에 의하여 제시된 기준 배합 비율표와 다른 비율로 배합하여야 올바른 조색이 이루어진다.

본 실험에서는 다양한 코팅제의 실제 배합 데이터 셋 중에서 다섯 종류의 코팅제를 표본으로 하여 실험을 수행한다. 코팅제 배합에 투입되는 정확한 재료명과 품목코드는 데이터 셋을 제공한 업체의 요청에 따라 대체코드로 표시하였다. 각 재료에 대한 대체코드는 도료의 기본구성(용제, 수지, 조색제, 첨가제)의 4종류로 분류하고 각 분류에 따라 일련번호를 적용하여 대체코드를 생성하였다. 각 대체코드의 구성은 Table 2와 같다.

Table 2. The Configuration of Alternative code for the Material Mixing of the Coatings

Classification code	Classification symbols	Applied to the sample sequence number
Solvents	A	1~4
Resin	B	1~4
Concentrates	C	1~9
Additives	D	1~14

3.2 데이터 최적화를 위한 실험계획 설계

일반적으로 실험계획법을 적용한 실험은 Fig. 6의 실험계획법 적용 순서도와 같은 순서로 수행된다[19].

① 실험목적의 설정

LOT별 코팅제 생산을 위해 제시되는 기준 배합 비율표에 따른 지시량과 불출량에 대하여 각 혼합재료의 품목코드당 발생하는 오차를 보정하기 위한 최적의 보정방정식과 계수를 도출한다.

② 특성치의 선택

코팅제 생산의 지시량과 불출량에 대하여 용제, 수지, 조색제, 첨가제 등 혼합재료 배합에 대한 상한값과 하한값 등의 특성값을 선택한다. 지시량, 불출량, 혼합재료 중의 조색 배합 시 결정요인이 되는 인자(수지, 조색제) 등을 고려하여 계측자 간의 측도와 계측기기의 측정 등을 통한 재현성과 안정성을 검토한다.

③ 인자와 수준의 선택

기준 배합 비율표에 따른 지시량과 불출량에 대하여 분석하고자 하는 인자를 선별한다. 인자는 가능한 한

서로 독립적이라고 생각되는 것을 선택하며, 인자의 적절한 배분을 통하여 최적의 실험을 계획하고 진행한다. 또한 지나치게 어려운 조건은 수준에서 제외시키고, 최적이라고 예상되는 조건을 선택하여 적용한다.

④ 실험계획법의 선정

실험 대상의 특징에 따라 적합한 실험계획법을 선정한다. 본 실험은 코팅제의 최적의 배합을 위한 계수와 방정식을 찾아내는 것이 실험목적이므로 혼합물 실험 계획법과 이원배치법을 선정한다. 또한, CCM을 통해 판별된 색상값을 활용하여 오차를 줄이기 위한 조색의 방향성 분석을 고려한다.

3.3 빅데이터의 적용을 위한 시스템 구성

실험계획법을 통하여 최적의 보정방정식과 계수를 도출한 후, 실험 결과를 이용하여 빅데이터 분석을 위한 기술구조(TA, Technical Architecture)를 설계한다.

제조현장에서는 계속적인 작업을 통하여 많은 양의 데이터가 수집되므로, 이에 대한 대응과 수집된 데이터의 분석 및 활용이 필요하다. 각 수집 정보는 전 공정단계에 대하여 매우 짧은 시간 단위로 수집되므로, 그 규모가 매우 크고 데이터의 변화가 빠르며 공정단계에 따라 다양한 형태의 데이터가 수집된다. 따라서 본 과제에서는 이러한 데이터의 분석을 위하여 빅데이터 분석 기법을 도입, 적용하고자 한다. 수집되는 데이터는 빅데이터의 조건인 3V를 충분히 만족한다고 판단되므로 빅데이터 분석 기법은 적절하다고 평가된다. Fig. 7은 빅데이터의 적용을 위한 코팅제 자동 배합 시스템의 데이터 흐름도이다.

Fig. 7의 시스템을 구성하기 위하여 작업공간의 환경에 대한 실시간 데이터 수집 장비가 필요하다. 본 논문에서 제안하는 데이터 최적화 모형에서는 사물 인터넷(Internet of Thing, IoT) 기술을 활용할 수 있는 데이터 수집 장비를 활용한다. IoT는 인간과 사물, 서비스 세 가지 분산된 환경 요소에 대해 인간의 명시적 개입 없이 상호 협력적으로 센싱, 네트워킹, 정보 처리 등 지능적 관계를 형성하는 사물 공간 연결망을 의미한다. IoT의 기반 기술로는 주위의 장비와 환경으로부터 정보를 수집하는 센싱 기술, 사물을 인터넷에 연결해주는 유무선 통신/네트워크 인프라 기술, 다양한 서비스와 형태에 적합하게 정보를 가공, 처리하고 기술을 융합하도록 해주는 서비스 인터페이스 기술을 들 수 있다.

제안하는 시스템에서는 이러한 작업환경의 정보를 수집하기 위하여 직접 개발한 IoT 데이터 수집 디바이스를 활용한다. 개발한 IoT 데이터 수집 디바이스는 온도, 습도, 전압, 전류 등의 환경 정보를 수집 및 측정하기 위한 센서 모듈과 입력 데이터의 가공을 수행하는 데이터 처리 모듈, 디지털 및 아날로그 데이터의 출력을 수행하는 모듈로 구성된다. Fig. 8은 개발하는 IoT 데이터 수집 디바이스의 구조를 보여준다. Fig. 9는 코팅제 배합/제조 장비와 개발 디바이스의 접속 방식을 보여준다.

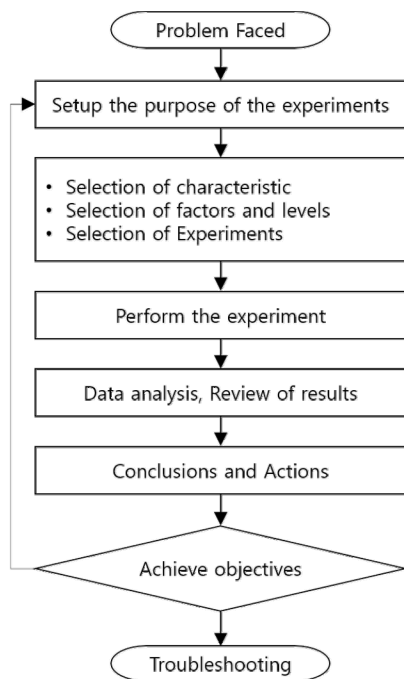


Fig. 6. The Flowchart for Applying to Design of Experiments

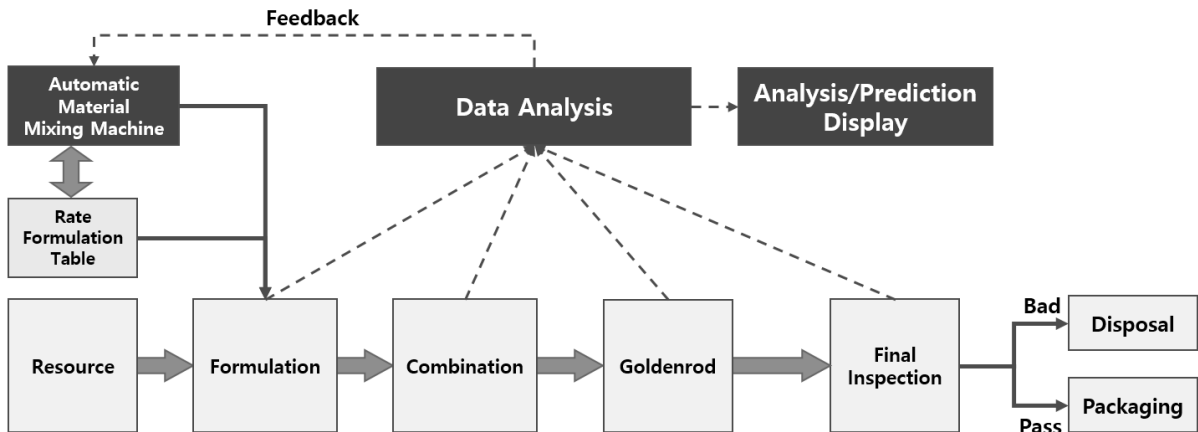


Fig. 7. The Data Flow of the Automatic Coatings Mixing System for the Big Data Analysis

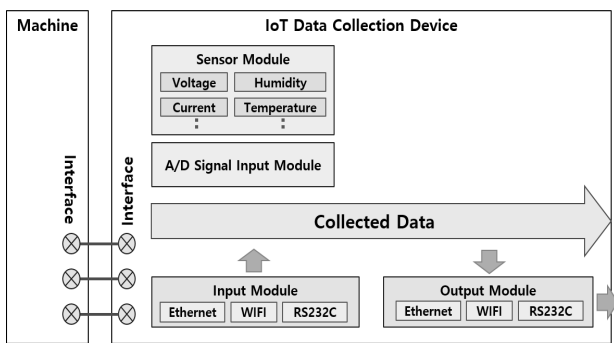


Fig. 8. The Developing IoT Data Collection Device

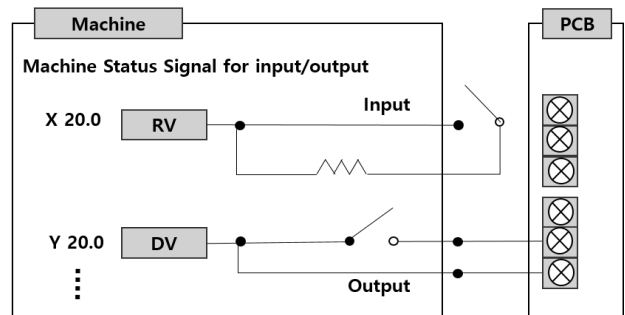


Fig. 9. The Connection Method for the Automatic Coatings Mixing System and Developing Device

수집된 데이터는 분석 과정을 거쳐 제조 공정에 피드백 됨으로써 그 효과를 확인할 수 있다. 분석 결과의 정확성 확보를 위하여, 최근 빅데이터의 분석에 많이 활용되고 보편화되고 있으며 이미 그 성능이 검증된 R 통계분석 툴을 기반으로 분석 결과를 검증한다. 실험을 통하여 도출된 보정방정식과 계수를 기반으로 구성된 시스템의 처리 결과와 실제 작업 중 수집되는 데이터 간의 오차의 계산과 피드백을 위한 데이터를 생성하기 위하여 R 통계분석 툴을 기반으로 검증된 분석 모델을 적용하고, 이를 통해 개발된 TA의 분석 결과는 제조 공정으로 피드백 되어 제조 공정의 개선에 활용한다.

#### 4. 모형 검증 및 분석

본 논문에서는 코팅제 제조 시 적용되는 배합 비율표를 기준으로 정확한 조색을 위하여 보정한 실제 배합 데이터를 이용하여 실험을 수행하였다. 실험 데이터는 제조현장에서 직접 측정된 실제 데이터를 적용하고 있으며, 데이터 셋을 제공한 업체의 현장 상황에 따라 제조현장의 온도, 습도 등 기초적인 환경데이터가 포함되어있지 않다. 그러나 실제 재료의 배합 시에는 환경적인 요인도 포함되므로 이러한 환경 정보도 함께 수집하여 처리할 필요가 있다. 따라서 실험 결과로 도출한 계수와 보정방정식은 환경 정보를 적용하여 계속적으로 갱신해나가야 하며, 제안하는 빅데이터 적용을 위한 시스템을 도입함으로써 개선이 가능하다.

Table 3. Some of the experimental data of the Coatings #01

	A4	B4	C1	C2	C4	C8	D3	D7	D14	TOT	L	A	B	ΔE
1	7.298	51.976	42.72	2.136	3.56	58.74	6.23	1.78	3.56	178	-1.51	3.96	-0.16	4.26
2	7.298	51.976	47.32	2.136	3.56	66.34	7.23	1.78	3.56	191.2	-0.9	3.06	0.45	3.22
3	7.298	51.976	52.32	2.136	3.56	66.34	8.93	1.78	3.56	197.9	-1.1	0.42	-0.82	1.43
4	7.298	51.976	57.32	2.136	3.56	73.34	9.63	1.78	3.56	210.6	-0.52	-0.22	-0.39	0.69
5	7.298	51.976	60.72	2.136	3.56	76.34	9.83	1.78	3.76	217.4	-0.1	-0.03	-0.2	0.22

Table 3은 실험에 사용된 데이터 중 코팅제 #01에 대한 표본 데이터의 일부이다. Table 3에서 A4, B4, C1, C2, C4, C8, D3, D7, D14는 코팅제 #01에 투입되는 재료이며, 배합 결과는 CCM을 통한 측정된 후 L, A, B, ΔE로 나타난다. Table 4의 기준에 따라 색상의 수치값으로 배합에 따른 조색의 방향성을 확인할 수 있다. Table 5는 각 결과값(ΔE, L, A, B)에 대한 회귀계수이다. A4, B4, C2, D7, D14 데이터는 수정량이 없으므로 아무런 관계성을 갖지 않는다. 구하고자 하는 결과값(ΔE, L, A, B)에 따라 해당하는 관계식을 적용하여 원하는 값을 얻을 수 있다.

Table 4. Direction and Meaning of the Mixing Color Data

	L	A	B
PLUS	White	Red	Yellow
MINUS	Black	Green	Blue

Table 5. The Regression Coefficients for the Mixing Ratio Change of the Coatings #01

	ΔE	L	A	B
Y	15.14626944	-9.332076451	1.336985761	-8.136271632
A4	0	0	0	0
B4	0	0	0	0
C1	-0.010514786	-0.005553122	-0.109200438	-0.210131435
C2	0	0	0	0
C4	-1.209200438	0.841391019	2.391949617	1.284884995
C8	0.003997809	0.096955093	0.109748083	0.224424973
D3	-1.022015334	-0.101314348	-1.231763417	-0.129025192
D7	0	0	0	0
D14	0	0	0	0

CCM 측정결과에 따른 ΔE에 대한 회귀계수를 계산하면 -0.010514786(C1), -1.209200438(C4), 0.003997809(C8), -1.022015334(D3) 이고 Y절편은 15.14626944이므로 ΔE에 대한 회귀방정식으로 식 (2)를 구할 수 있다.

$$\Delta E = 15.14626944 - 0.010514786 \times C_1 - 1.209200438 \times C_4 + 0.003997809 \times C_8 - 1.022015334 \times D_3 \quad (2)$$

같은 방법에 따라 L, A, B에 대한 식 (3), 식 (4), 식 (5)를 구할 수 있다.

실험 데이터에서는 4개의 재료만 수정되었고, Table 5에서 구할 수 있는 관계식 또한 4개이므로 식 (2)~식 (5)를 이용함으로써 연립방정식 계산을 통하여 단 하나의 답을 찾을 수 있다. 최종 계산된 C1, C4, C8, D3의 값을 비율로 변환하여 배합 공정에 적용할 경우 정확한 결과를 얻을 수 있다고 예상할 수 있다.

$$L = -9.332076451 - 0.005553122 \times C_1 + 0.841391019 \times C_4 + 0.096955093 \times C_8 - 0.101314348 \times D_3 \quad (3)$$

$$A = 1.336985761 - 0.109200438 \times C_1 + 2.391949617 \times C_4 + 0.109748083 \times C_8 - 1.231763417 \times D_3 \quad (4)$$

$$B = -8.136271632 - 0.210131435 \times C_1 + 1.284884995 \times C_4 + 0.224424973 \times C_8 - 0.129025192 \times D_3 \quad (5)$$

코팅제 #01의 배합 데이터와 계산된 회귀방정식을 이용하여 혼합물 실험계획법을 적용하면 Fig. 10~Fig. 15의 그래프를 얻을 수 있다. Fig. 10~Fig. 15의 그래프는 각각 실질적인 실험 데이터를 적용한 경우와 데이터의 변화에 따른 결과값 변화를 실험하기 위한 가상 데이터를 포함한 경우에 대하여 각각 ΔE의 혼합물 등고선도, Cox 반응 궤적도, 최적화에 대한 비교 내용을 보여준다.

실험을 위한 실험점의 개수는 식 (6)에 따라 구할 수 있다.

$$n = k + \frac{k(k+1)}{2}, \quad (k: \text{인자의 개수}) \quad (6)$$

그러나 Fig. 10과 Fig. 11에서 나타나고 있는 기존 데이터를 적용한 실험의 분석 결과와 가상 데이터를 포함한 실험 데이터의 분석결과에 대한 ΔE의 혼합물 등고선도 행렬의 비교결과에서 알 수 있듯이 지시량이 정해진 상태에서는 인자 개수에 대한 실험점의 개수가 최소 요구개수보다 적은 경우에도 유사한 결과를 얻을 수 있음을 확인할 수 있었다. 그리고 전체적으로 미치는 영향을 분석할 때, 배합 시 C1(흰색을 표현하는 조색제)에 대한 의존도가 높고, 또한 C8(흰색 이외의 색상을 표현하는 조색제의 하나)은 다른 성분에 비해 배합 결과에 영향을 크게 미치지 않는다는 분석 결과가 나왔다. 그러나 Cox 반응 궤적도 및 최적화 그림의 경우에는 입력된 자료에 대한 결과값의 추출이 용이하지 않았다.

Fig. 12와 Fig. 13에서는 기존 데이터를 이용한 실험과 가상 데이터를 포함한 실험에서의 Cox 회귀모형 분석에 따른 Cox 반응 궤적도를 비교하고 있다. Cox 회귀모형 분석은 시간-사건 데이터의 예측 모형 작성을 위해 사용되는 회귀 분석 방법이다. 주어진 예측변수에 대하여 특정 시간(t)에 중요한 사건이 발생할 확률을 예측하는 생존함수를 생성하고, 생성된 함수를 통해 예측변수의 측정값을 갖고 있는 새로운 케이스에 모형을 적용할 수 있다. 생존함수의 모양과 예측변수의 회귀계수는 관측된 개체를 통해 추정된다. 기존 데이터만을 사용한 실험 결과인 Fig. 12는 단순한 선형성만을 표현하고 있다. 그러나 실제 배합을 수행할 경우, 결과는 단순한 선형성만을 가지고 있지 않았다. 반면, 실험을 위한 가상 데이터를 포함한 Fig. 13에서 나타난 Cox 반응 궤적에서는 실제 배합의 경우와 유사한 결과를 도출할 수 있었다.



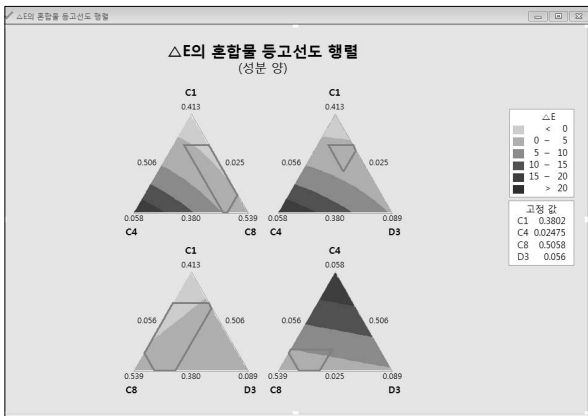


Fig. 10. The Contour Matrix for the Mixing Material of the  $\Delta E$  for the Existing Data

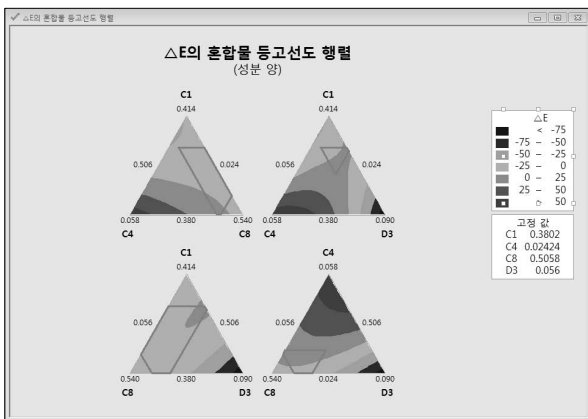


Fig. 11. The Contour Matrix for the Mixing Material of the  $\Delta E$  with Virtual Data

Fig. 14와 Fig. 15에서 볼 수 있는 기존 데이터 실험 경우와 가상 데이터를 포함한 실험 데이터 경우에서의 최적화 그림의 비교결과에서 C1, C8은 목표값 및 최적화 지점에 대한 반응을 보여주고 있다.

Fig. 14~Fig. 15의 기존 데이터 실험 경우와 가상 데이터를 포함한 실험 데이터 경우에 대한 비교 그래프에서 알 수 있듯이 조색작업에 가장 영향을 미치는 것은 C1, C8로 알 수 있으며, 실질적인 배합의 중요 원료는 C1인 것을 최적화 그림에서 알 수 있다.

### 5. 결론

본 논문에서는 사물 인터넷 기술과 빅데이터 분석을 활용한 실험계획법 기반의 코팅제 배합 데이터에 대한 최적화 모형을 제안하였다. 실험을 통하여 실제 생산현장에서 사용하는 코팅제 배합 기준 데이터와 발생한 오차에 대하여 작업자가 수정한 보정 결과 데이터를 실험계획법을 기반으로 분석하여 최적의 코팅제 배합 기준 데이터를 계산하였다. 또한 빅데이터 분석 기술과 사물 인터넷 기술을 활용하여 기존의

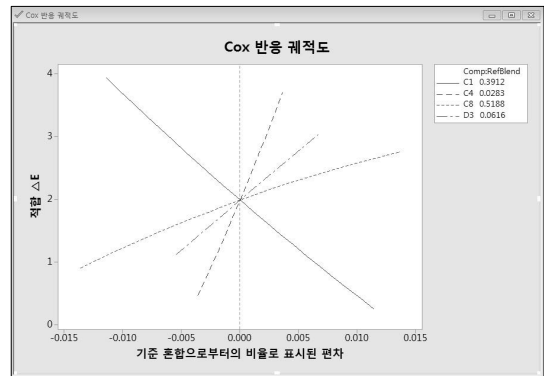


Fig. 12. The Cox Reaction Trajectory for the Existing Data

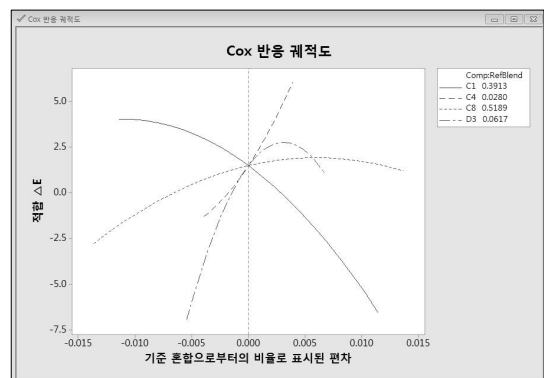


Fig. 13. he Cox Reaction Trajectory with Virtual Data

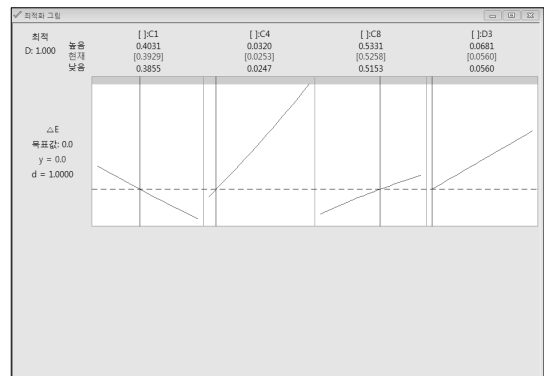


Fig. 14. The Optimizing Pictures for the Existing Data

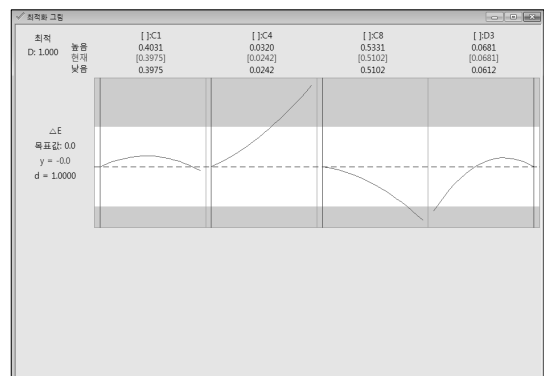


Fig. 15. The Optimizing Pictures with Virtual Data

코팅제 배합 기준 데이터만을 적용한 공정이 아니라 제조 환경 정보와 상황 정보를 이용하여 색상과 품질 유지에 가장 중요한 인자를 검색하고 기준값을 보정하는 최적화 모형을 도출하였다. 다음으로 본 논문에서 도출한 코팅제 배합 비율 데이터 최적화 모형에서 코팅제 배합비율 데이터 최적화 모형에서 원료코드별 지시량과 불출량, 도료의 기본색의 영향, 조색 시 조색제와 수지의 중요도에 대하여 실험계획법의 하나인 혼합물 실험계획법을 통해 코팅제 조색결과의 혼합물 등고선도, 최적화, Cox 반응 궤적도를 구하고 조색 데이터의 방향성과 배합의 연관성, 결과물과 목표 간의 오차 감소 방향 등을 분석하였다. 실험 및 분석을 통해 확보된 기준 데이터는 제조 공정에 적용할 경우 배합의 정확도 향상과 LOT별 작업시간 단축을 가능하게 해주고, 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.

향후 연구 분야는 축적된 데이터를 개발된 서비스 플랫폼(빅데이터 - R 통계)을 바탕으로 상황에 맞는 TA를 설계하고, 수집된 자료를 데이터 검증 분석 방법에 적용하여 표준 아키텍처를 제시 및 객관적인 검증을 할 수 있는 기술 연구에 중점을 두고자 한다.

### References

[1] Korea Institute of Science and Technology Information, Functional Inorganic Coatings, [Internet], [http://mirian.kisti.re.kr/publication/view.jsp?record\\_no=155&cont\\_cd=KI](http://mirian.kisti.re.kr/publication/view.jsp?record_no=155&cont_cd=KI)

[2] Analyst Anish Gaddam interviewed by Sue Bushell in Computerworld, on 24 July 2000 (“M-commerce key to ubiquitous internet”).

[3] Ashton Kevin, *That ‘Internet of Things’ Thing, in the real world things matter more than ideas*, RFID Journal, 2009.

[4] Douglas Laney, “3D Data Management: Controlling Data Volume, Velocity and Variety”, *META Group*, Feb., 2001.

[5] Mark Beyer, Gartner Says Solving ‘Big Data’ Challenge Involves More Than Just Managing Volumes of Data, Gartner. Retrieved 13 July 2011. [Internet], <http://www.gartner.com/it/page.jsp?id=1731916>.

[6] Hong Keun Yoon, “Research on the Application Methods of Big Data within the Cultural Industry”, *Global Cultural Contents*, Vol.10, pp.163-166.

[7] Dhruva Borthakur, Apache Hadoop [Internet], <http://hadoop.apache.org/>

[8] Apache Flume [Internet], <http://flume.apache.org>

[9] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2008.

[10] Heewon Jeon, “Big Data Analysis with R: R Distributed Programming”, *Microsoft*, pp.186-193, Sep., 2011.

[11] Lee, Kae Gil, “*Statistical analysis using the MATLAB*”, Ajin, Seoul, 2013.

[12] Hyeyoung Park, Kwanyong Lee, “Pattern Recognition and Machine Learning: From the ground up to take advantage of”, *Ehan Media*, Goyang, 2011.

[13] Jung, D. B. and Won, T. Y, Time series data and analysis of simplification I, *Hannarae Publishing Group*, Seoul, 2001.

[14] Kyung-Sook Woo and Young-Jeon Shin, “A systematic review of studies using time series analysis of health and welfare in Korea”, *Korean Data and Information Science Society*, Vol.25, No.3, pp.579-599, 2014.

[15] Min-Hoi Koo, Jin-Woo Lee and Hoe-Cheol Cha, “CCM System Practical Technology Research for Laboratory Dyeing Machine”, *Korean Institute of Industrial Technology*, 1993.

[16] Stanley, J. C., “The Influence of Fisher’s ‘The Design of Experiments’ on Educational Research Thirty Years Later”, *American Educational Research Journal* 3(3), pp.223-229, 1966.

[17] Box, JF., “R. A. Fisher and the Design of Experiments, 1922-1926”. *The American Statistician* 34(1), pp.1-7, Feb., 1980.

[18] Yates, F., “Sir Ronald Fisher and the Design of Experiments”. *Biometrics* 20(2), pp.307-321, June, 1964.

[19] Sung-Hyun Park, Jong Wook Kim, “*Modern Design of Experiments using MINITAB*”, Minyoungsa, Seoul, 2011.



### 노성여

e-mail : shtjddu@paran.com

1998년 부경대학교 (전산정보)대학원(석사)

2014년 부경대학교 (기술경영)대학원 박사

수료

현 재 부경대학교 기술경영협동과정 박사

과정

관심분야: 향만물류 시스템, 데이터 분석, 프레임워크



### 김영진

e-mail : youngk@pknu.ac.kr

1987년 한국과학기술원 경영과학(이학사)

1992년 한국과학기술원 경영과학(공학석사)

2000년 Ph.D in Industrial Engineering,

Clemson University

현 재 부경대학교 시스템경영공학부 교수

관심분야: 응용통계, 품질공학, 최적화 기법의 응용, 경제성 분석