

Information Visualization for the Manufacturing Process Optimization Based on Design of Experiment and Data Analysis

Kim Jae Chun[†] · Jin Seon A^{**} · Park Young Hee^{***} · Noh Seong Yeo^{****} · Lee Hyun Dong^{*****}

ABSTRACT

Data visualization technology helps people easily understand various data and its analysis result, so usefulness of it is expected in the real industrial manufacturing sites. The large amount of data which is occurred at the manufacturing sites is able to fulfill very important roll to improve the manufacturing process. In this paper, we propose an information visualization for the manufacturing process optimization based on design of experimental and data analysis. The manufacturing process may be improved and be reduced cause of faulty by providing the easy-process analysis to understand the operation site through the information visualization of data analysis result.

Keywords : Information Visualization, Manufacturing Process Optimization, Design of Experiment, Data Analysis

실험계획법과 데이터 분석 기반의 제조공정 최적화를 위한 정보 시각화

김재천[†] · 진선아^{**} · 박영희^{***} · 노성여^{****} · 이현동^{*****}

요 약

데이터 시각화 기술은 다양한 데이터와 그 분석 결과를 쉽게 이해할 수 있도록 도와줌으로써 제조현장과 같은 실제 산업현장에서도 그 유용성이 기대되고 있다. 제조현장에서 발생하는 대량의 데이터는 제조 기술의 표준화를 위한 기반 데이터가 될 수 있으며 제조공정의 개선을 위하여 매우 중요한 역할을 수행할 수 있다. 본 논문에서는 실험계획법과 데이터 분석 기반의 제조공정 최적화를 위한 정보 시각화 방법을 제안한다. 데이터 분석 결과의 정보 시각화를 통하여 작업 현장에 이해하기 쉬운 분석 결과를 제공함으로써 다양한 불량원인을 감소시키고 제조공정을 개선시킬 수 있다.

키워드 : 정보 시각화, 제조공정 최적화, 실험계획법, 데이터 분석

1. 서 론

국내 제조업 시장은 빠르게 변화하는 제조환경의 변화에 대응하기 위하여 다양한 노력과 기술 개발을 시도하고 있으나 업체의 영세성 등 구조적, 기술적인 취약성으로 인해 어려움을 겪고 있다. 특히, 충분한 기술 및 자금력을 가진 대

기업과 그 1차 협력사와는 달리, 국내 제조업체의 대부분을 차지하는 2, 3차 협력사의 경우 업체의 규모 및 장비의 영세성 때문에 작업자의 수작업과 육안 검사에 의해 품질 검사를 진행하는 경우가 대부분이다. 또한, 제조환경 및 공정 데이터에 대한 객관적인 분석이 지원되지 않아 숙련자의 공백이 발생할 경우 비숙련자에 의한 인적 오류의 발생률도 증가하게 된다.

따라서 제조업체의 품질관리 및 공정 개선을 위해서는 제조공정에 대한 데이터의 분석과 검증, 예측 기술을 적용한 정보의 제공과 생산된 정보를 쉽게 이해하고 적용할 수 있도록 지원하는 데이터 시각화 기술은 제조환경 및 공정 개선에 매우 중요하다.

사출 및 성형공정에서는 다양한 혼합물을 원자재로 하여 공정을 수행한다. 제품의 품질 향상을 위하여 원자재가 되

※ 본 논문은 중소기업청에서 지원하는 2014년도 산학협력 기술개발사업 (No. C0219239)의 연구수행 결과물임.

† 준 회 원 : ㈜나라시스템 대표이사

** 정 회 원 : ㈜나라시스템 기업부설연구소 선임연구원

*** 비 회 원 : 한국폴리텍대학 정보통신시스템과 조교수

**** 정 회 원 : 동명대학교 항만물류시스템학과 교수

***** 정 회 원 : ㈜나라시스템 기업부설연구소 연구소장

Manuscript Received : July 31, 2015

First Revision : September 1, 2015

Accepted : September 3, 2015

* Corresponding Author: Hyundong Lee(win4class@hanmail.net)

는 혼합물에 대한 배합 정보의 최적화는 매우 중요하다. 그러나 국내 중소 제조업체에서는 여전히 작업자의 경험과 수작업에 의존하여 배합공정을 수행하는 실정이며 효율적인 배합공정을 위해 보다 지능화되고 자동화된 처리 기술의 필요성이 대두되고 있다[1].

본 논문에서 제안하는 프레임워크를 이용하여 혼합물의 배합 공정에서 발생한 데이터를 분석하고 분석 결과를 이용하여 최적화된 배합 비율 데이터를 계산한 후, 이해하기 쉬운 시각화 기술을 적용하여 최적화된 배합 기준 비율을 작업자에게 제시함으로써 제조 환경에 따라 보정되어야 하는 정확한 배합 표준 데이터의 부재로 인해 발생하는 품질 저하와 원가 손실을 감소시키고, 배합 공정에서 발생한 오차 데이터를 분석하여 관계식을 도출함으로써 차후 발생 가능한 오차를 감소시키기 위한 정보를 제조현장에 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 본 논문과 관련된 연구와 배경 지식을 소개하고, 3절에서는 데이터 분석 기반의 프레임워크를 소개한다. 4절에서는 제시한 프레임워크를 이용하여 실제 배합작업 데이터를 분석하고 최적 값을 도출하고 도출된 분석 결과에 프레임워크의 시각화 기능을 적용한다. 마지막으로 5절에서는 결론과 향후 연구에 대하여 논의한다.

2. 관련 연구

2.1 실험계획법(Design of Experiments, DoE)

실험계획법이란 어떤 방식으로 실험을 수행하고, 데이터를 수집하며, 어떤 통계적 방법을 적용하여 데이터를 분석하면 최소의 실험으로 최선의 결과를 얻을 수 있는가를 계획하는 방법이다[2].

실험계획법은 랜덤화, 반복, 블록화로 구성된다. 랜덤화는 선택되지 않은 요인이 편향된 실험 결과를 만드는 것을 막기 위하여 사용하는 방법이다. 반복은 동일한 조건에서 실험을 두 번 이상 수행하는 것이다. 블록화는 실험을 가능한 한 동질적인 부분으로 나누어 여러 블록으로 구성한 후, 각 블록별로 인자의 효과를 조사하는 것이다.

실험계획법에는 요인 배치법, 분할법, 반응표면 계획법, 혼합물 실험계획법, 로버스트 실험계획법 등이 있으며 제조업체의 원재료의 특징 분석을 위하여 수행연구에서는 요인 배치법의 이원 배치법과 혼합물 실험계획법을 실험에 적용하였다. 본 논문에서 제안하는 방법에서는 프레임워크의 분석 및 검증 컨테이너에서 실험계획법과 기계학습을 포함한 데이터마이닝 기법을 적용한다.

2.2 데이터 분석

데이터란 변수의 질적 또는 양적 특성을 나타내는 변수 또는 정보 그룹의 세트를 의미한다. 이러한 데이터가 가지는 정보를 다양한 방법으로 분석하여 유용한 의미를 밝혀내는 활동을 데이터 분석이라고 하며 데이터 처리 및 분석을 통한 데이터의 활용은 생산성의 향상과 고부가가치 및 고품

의 창출 등 경제적 가치 창출을 위한 핵심기술로 평가받고 있다. 기업체들은 자사의 경영 전략에 데이터 분석을 도입하여 수익 증대를 실현할 수 있으며, 데이터 분석을 공공영역에 도입할 경우에는 높은 사회적·경제적 효과가 발생할 것으로 기대되고 있다.

데이터 분석은 통계학을 기반으로 하는 분석 기법과 데이터마이닝을 기반으로 하는 분석 기법이 있으며, 최근에는 빅데이터 분석을 대상으로 하는 데이터마이닝 기반의 분석 기법이 주목받고 있다. 대표적인 분석 모델로는 선형회귀분석, 군집분석, 시계열분석 등을 들 수 있다.

선형회귀분석은 선형성을 가지는 독립변수와 종속변수의 관계를 설명하거나 두 변수의 관계성을 이용하여 데이터의 변화를 예측하는 통계방법이다. 즉 회귀분석에서 독립변수에 따라 종속변수의 값이 일정한 패턴으로 변해가는데, 이러한 변수 간의 관계를 나타내는 회귀선이 직선에 가깝게 나타나는 경우를 선형회귀분석이라고 한다[3]. 독립변수가 하나인 경우를 단순회귀분석, 여러 개인 경우를 다중회귀분석이라고 한다.

군집분석이란 데이터의 군집을 찾는 방법이다. 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, 요인 분석 등으로 미리 군집을 예상하여 합쳐가는 비계층적 방법도 있다[4]. 군집을 찾아내기 위해서는 각각의 군집에 개별 데이터를 할당하여야 하는데 어떤 할당 방법이 최선인지 알 수 있는 정의된 방법은 아직까지 존재하지 않는다. 이에 이러한 데이터의 할당은 다음의 3가지 방법을 주로 사용한다[5].

- kmeans와 같은 알고리즘을 사용하여 사용자가 군집의 수를 설정한 후 분할하는 방법
- 분리된 개별 개체를 클러스터링 기법을 사용하여 하나로 묶는 계층적 방법
- 모든 개체를 하나로 묶은 후 이를 모든 개체가 서로 다른 군집으로 분리될 때까지 나누는 방법

시계열 데이터는 어떤 현상에 대하여 시간의 흐름에 따라 일정한 간격으로 관측하여 기록한 데이터를 의미한다. 일반적으로 동일한 시간 구간으로 나누어진 숫자들로 이루어진 벡터의 형태를 가진다[5]. 시계열 분석이란 시계열 데이터에 바탕을 둔 분석방법으로, 시계열 데이터가 시간에 따라 변화하는 구조 사이에 존재하는 다양한 확률적, 통계적 성질을 파악하고, 이를 바탕으로 확률 현상의 미래 형태를 예측하는 통계분석을 의미한다[6]. 시계열 분석은 트렌드, 연속 의존성, 정상성의 3가지 중요한 개념을 기반으로 한다. 3가지 개념이 가지는 의미는 다음과 같다.

- 대부분의 시계열 분석은 데이터가 트렌드를 가지지 않는다고 가정하며, 데이터가 일정하게 상승 또는 하강하는 트렌드를 가진 경우, 분석 전에 트렌드 제거를 할 수 있다.
- 연속 의존성은 시계열 데이터의 인접한 값 사이에 상관

관계가 있을 경우 나타난다.

- 정상성은 기술적인 개념이며 시계열을 어디에서부터 살펴보다라도 같은 성질을 가지는 경우를 말한다.

본 논문에서는 이상과 같은 다양한 데이터 분석 기법을 기반으로 개발된 프레임워크를 제안하고 제시된 프레임워크를 이용하여 코팅제 배합공정을 분석한다.

2.3 데이터마이닝(Data Mining)

데이터마이닝이란 대용량의 데이터에서 숨어있는 유용한 상관관계와 의미 있는 패턴을 파악하여 데이터를 자동으로 분석, 분류함으로써 미래에 실행 가능한 정보를 추출하고, 예측을 수행하거나 의사결정에 활용하는 기술을 말한다. 데이터마이닝에는 통계 모델링을 기반으로 하는 방법과 기계 학습을 기반으로 하는 방법이 있다. 데이터마이닝이 통계분석과 다른 점은 어떤 가설이나 가정에 따른 분석이나 검증이 아니라 존재하는 데이터로부터 유용한 패턴을 찾는 기술이라는 점이다. 데이터마이닝을 구성하는 요소들은 다음과 같다.

- 지식발견(KDD, Knowledge Discovery in Databases) : 지식을 추출하는 전 과정을 의미한다. 데이터의 정제, 통합, 선택, 변환, 마이닝, 패턴평가, 지식 프레젠테이션의 과정을 가지며 데이터 정제부터 변환까지의 과정은 데이터마이닝을 위한 전처리 과정으로 볼 수 있다[7].
- 기계학습(Machine learning) : 데이터를 지능 행위로 변환하는 컴퓨터 알고리즘을 연구하는 인공지능의 한 분야이며 자동적인 학습 기법을 설계 및 구현하는 기술이다[8].
- 패턴인식(Pattern recognition) : 대용량의 데이터를 분석하여 숨어있는 유용한 패턴을 찾아내는 다양한 기법들을 말한다. 데이터수집, 특징선택, 모델선택, 훈련, 평가의 과정을 거친다[9].
- 통계학(Statistics) : 주어진 자료에서 합계나 평균과 같은 필요한 정보를 수집, 정리, 요약하거나, 표본 자료에서 얻은 정보를 이용하여 모집단에 대한 정보를 예측하고 불확실한 사실에 대한 결론을 이끌어내는 학문으로 데이터마이닝의 대부분은 통계학의 이론을 활용한다.

본 논문에서는 기업의 경영자 및 제조현장의 작업자에게 제공하기 위하여 각 데이터에 대한 관계성 분석 결과를 데이터마이닝과 기계학습을 활용하는 프레임워크를 통하여 제공한다.

2.4 R 기반 통계분석

R은 오픈소스 프로젝트의 하나로 통계 분석과 시각화에 효과적인 분석도구이다. 데이터 분석을 위한 입출력과 핸들링, 분석, 그래픽 등 최신 알고리즘과 라이브러리를 제공하며 수많은 사용자에게 의하여 충분히 검증된 통계 분석 프로그램이다. In-Memory 컴퓨팅 기술을 통한 빠른 처리속도와 분석 및 시각화를 위한 강력한 그래픽 함수를 지원한다.

본 논문에서 제시하는 프레임워크는 R을 활용하여 다양한 데이터를 분석하고 해당 데이터의 신뢰성을 검증하는 API 셋을 개발한다. 개발하는 API 셋은 본 논문에서 제시하는 프레임워크의 구성요소의 하나인 분석 및 검증 컨테이너(Analysis & Verification Container)에 해당한다.

2.5 정보 시각화(Information Visualization)

정보 시각화라는 용어는 1999년 Stuart K. Card가 처음 사용한 것으로, ‘컴퓨터 기반이며 상호작용적이고 인지의 확대를 위한 기초적 데이터의 시각적 표현을 사용한 것’이라고 정의하고 있으며[10] 메시지를 전달하기 위해서 작성하는 이미지나 다이어그램, 혹은 애니메이션들에 대한 기술로 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말한다[11]. 정보 시각화의 목적은 도표(graph)와 같은 그래픽 요소를 활용하여 정보 및 메시지를 명확하고 효과적으로 전달하는 것이다. 특히 데이터 시각화의 종류로서 공간자료의 시각화는 보다 효율적인 자료의 탐색과 분석을 위해 매우 중요하다. Table 1은 정보 시각화의 장점과 유의점을 보여준다.

Table 1. Advantage and Attention Point for the Information Visualization

Advantage	<ul style="list-style-type: none"> - Expansion of information processing capability - Differentiate a lot of data at the same time - Perceptual reasoning available - Emotional expression available - Access to a wide layer - Delivery of information in the context - Multi-dimensional information
Attention Point	<ul style="list-style-type: none"> - Requirements of intellectual ability to interpret process - Risk of distortion of information - Excessive degradation efficiency of visualization - Cultural factors involved

정보 시각화는 최근의 빅데이터 처리기술에서 크게 주목을 받고 있으며 데이터마이닝과 같은 전통적인 데이터 분석 기술과 함께 활용되고 있는 빅데이터 처리기법의 한 부분이다. 분석된 데이터를 사용자가 이해하기 쉽게 시각화 처리를 수행하고 있지만 아직까지 객관적 정확성이 부족한 편이다. 엑셀, 구글 차트 API 등을 활용하는 기초적인 수준에서부터 상호작용을 지원하는 GUI 컨트롤과 전문가 수준의 분석도구에 이르기까지 다양한 도구가 개발 및 활용되고 있다. 본 논문에서는 제안하는 프레임워크를 통하여 다양한 시각화 API 셋을 지원한다.

2.6 프레임워크

프레임워크란 프로그래밍에서 특정 운영체제를 위한 응용 프로그램의 표준 구조를 구현하는 클래스와 라이브러리의 모임이다. 프레임워크를 개발에 적용할 경우, 기반구조를 그대

로 사용하고 애플리케이션만의 비즈니스 로직과 관련된 부분만을 구현하면 되기 때문에 구현 양이 적어지게 되고 보다 빠르게 애플리케이션을 만들 수 있다. 또한 같은 프레임워크가 사용된 애플리케이션은 서로 비슷한 구조를 가지게 되어 관리가 용이해지며, 프레임워크를 재사용하게 되면 기존에 수행된 테스트들도 재사용하게 되는 것이라 애플리케이션의 테스트에 대한 부담을 감소시키고, 프레임워크를 사용하면서 새로 구축한 모듈에 대한 테스트만 수행해도 되므로 테스트에 대한 작업시간을 단축할 수 있는 장점을 가진다.

본 논문에서 제안하는 프레임워크는 데이터 분석을 기반으로 구성되며, IoT 기반의 데이터 수집 디바이스를 통한 환경정보, 작업자로부터 입력되는 작업정보, 원자재 배합을 위하여 투입되는 원자재별 투입량 정보를 입력받아 분석을 수행하고 그 분석 결과에 대한 다양한 시각화 기술을 적용한 표현 기능을 가진다.

3. 데이터 분석 기반 프레임워크 설계

3.1 데이터 분석 기술

본 논문에서는 혼합물 원자재 배합공정에서 발생하는 다양한 데이터를 분석하기 위하여 오픈소스 통계 분석 도구인 R을 활용하였고 실제 배합공정에서 수집한 데이터와 작업지시 데이터를 대상으로 분석 작업을 수행하였다. 그러나 실제 작업에서는 일반적으로 예측할 수 없는 다양한 환경을 구축할 수 없고, 원자재 투입 정보 또한 실제 제조를 위하여 주어진 배합 비율표에 따라 작업이 수행되므로 각 원자재 간의 관계성을 분석하기 위한 충분한 범위의 데이터를 수집할 수 없었다. 또한 이미 숙련된 작업자가 작업을 진행하고 있으므로 비숙련자에 의한 인적 오류의 발생 횟수가 현저히 낮아 데이터의 분석에 의하여 각 변수의 관계성을 구하기 어려웠다. 이러한 이유로 본 논문에서는 실험계획법을 도입하여 실험조건을 정의하고, 정의된 조건에 따른 데이터를 생성하여 분석 작업을 수행한 후 실제 배합공정 데이터의 결과와 비교함으로써 결과의 정확성을 검증하는 방법을 이용하였다.

실험계획법은 실험을 어떻게 수행하고, 데이터를 어떻게 수집하며, 데이터를 어떤 통계적 방법으로 어떻게 분석하면 최소의 실험만으로 최대의 정보를 얻을 수 있을지에 대하여 계획하는 방법을 말한다[2]. 본 논문의 실험에서는 원자재의 배합을 통해 제조되는 제품의 특징에 따라 혼합물 실험계획법을 적용하였다. 혼합물 실험계획법이란 여러 개의 성분의 혼합으로 이루어져 있고 각 성분의 혼합비율만이 문제가 되는 경우, 어떠한 성분이 관심이 있는 반응 값에 유의한 영향을 미치며, 반응을 최대 또는 최소로 만드는 최적 혼합비율이 무엇인지 찾고자 하는 실험계획을 말한다[2]. 혼합물 실험계획법을 통하여 생성된 공정 데이터는 회귀분석기법을 기준으로 분석 작업을 수행하였다.

3.2 데이터 분석을 위한 데이터 셋

본 논문에서 제시하는 프레임워크를 활용하여 데이터 분석 및 시각화를 수행하기 위해서는 데이터 셋이 필요하다. 실험에서는 원자재 배합공정에서 발생하는 실제 공정데이터를 활용하며, 해당 데이터에 대한 확보방안은 다음과 같다.

- 요구사항 분석 결과 및 수집 데이터 정의/분석을 통하여 제조현장에서 수집할 수 있는 데이터를 수집, 정리
- IoT 기반 스마트 디바이스를 활용하여 실시간으로 데이터 수집을 수행하며, 수집된 데이터는 분류모듈을 통하여 데이터 셋으로 가공, 활용
- 원자재 배합 기준 비율표와 특성값, 예외사항 등 수집 후 공정조건 점검일지 및 제조과정 히스토리 파일 수집
- 초기 허용 불량 수 관리 및 불량산출 근거에 대한 데이터의 비교 분석
- 작업자는 자주 검사 시트, Set-up 검사 기준서 등을 기준으로 불량사례를 찾은 후 IoT 스마트 디바이스에서 수집된 정보와 비교, 분석

수집된 데이터는 분석 과정을 거쳐 제조공정에 피드백 됨으로써 그 효과를 확인할 수 있다. 분석 결과의 정확성 확보를 위하여, 수많은 사용자를 가지며 그 성능이 이미 검증된 오픈소스 통계 분석 툴인 R을 기반으로 분석 결과를 검증하였다. 실험을 통하여 도출된 보정방정식과 계수를 기반으로 구성된 시스템의 처리 결과와 실제 작업 중 수집되는 데이터 간의 오차 계산과 피드백을 위한 데이터를 생성하기 위하여 R 통계분석 툴을 기반으로 검증된 분석 모델을 적용하고, 분석 결과를 제조공정으로 피드백 함으로써 제조공정의 개선에 활용한다. 제조 장비가 제공하는 데이터뿐만 아니라 온도, 습도 등의 환경측정 데이터와 작업지시내용을 데이터로 입력받아 분석에 활용한다. 제조 장비가 주변 환경에 대한 데이터를 측정하여 제공하는 경우는 제조 장비의 측정값과 개발하는 디바이스의 센서 모듈을 이용한 환경측정 데이터를 비교하여 장비와 환경 사이의 관계성을 분석할 수 있다. 제조 장비가 환경 데이터를 측정 및 제공하지 않는 경우, 디바이스의 자체 측정 데이터를 활용하여 환경과 제조 과정 사이의 관계성을 분석할 수 있다.

3.3 프레임워크 설계

본 논문에서는 혼합물 원자재의 배합공정을 분석하기 위한 데이터 분석 기법을 기반으로 하는 프레임워크를 개발, 활용하였다.

제안하는 프레임워크는 IoT 기반 센서 디바이스 제어 규칙 관리, 표준 데이터 셋 관리, 산업 유형별 규칙관리를 수행하는 IoT Device Container, Hadoop 기술을 기반으로 빅데이터 저장기술의 확장 모듈 및 데이터 변환을 수행하는 Big Data Container, 변환된 데이터에 대한 분석 및 검증, 예측, 시각화 기술을 관리하는 Analysis & Verification Container, 산업 유형별로 저장된 데이터 셋을 이용한 유형

별 제어 규칙을 관리하는 Industry Process Container로 구성한다.

데이터 수집을 위한 IoT 기반 스마트 디바이스를 통한 인터페이스를 제공하며 각 산업별로 특화된 데이터 및 다양한 환경, 작업 데이터를 입력받을 수 있다. 제안하는 프레임워크는 IoT 데이터 수집 인터페이스를 통하여 다양한 산업 현장에 적용 가능하다. Fig. 1은 개발하는 프레임워크의 구조를 나타낸다.

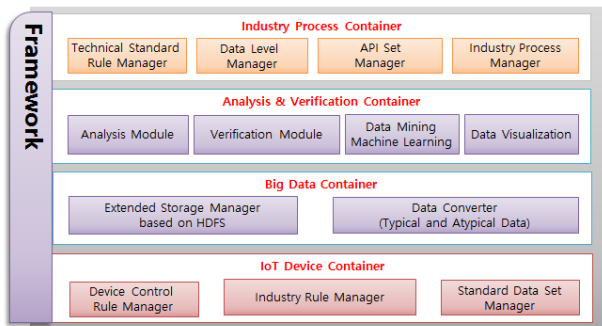


Fig 1. Architecture of Proposed Framework

IoT 디바이스 컨테이너는 제조 장비가 제공하는 데이터, 환경 데이터 및 작업 지시 데이터를 기준으로 수집한다.

저장된 표준 데이터를 이용하여 생성되는 디바이스 제어 규칙을 관리하는 Device Control Rule Manager, 산업 유형별로 저장된 데이터 셋을 이용한 유형별 제어 규칙을 관리하는 Industry Rule Manager, 각 부분에 대한 표준 데이터 셋을 관리하는 Standard Data Set Manager로 구성된다.

주기적인 측정 데이터 및 표준데이터에 대한 백업을 지원하며, IoT 데이터 수집 인터페이스를 통해 수집된 데이터가 어떤 변환과정을 거칠 것인지 자동으로 분류하여 빅데이터 컨테이너로 전달한다. Fig. 2는 IoT 디바이스 컨테이너의 구성을 보여준다.



Fig 2. IoT Device Container

빅데이터 컨테이너(Big Data Container)는 빅데이터에 대한 전처리를 수행하는 컨테이너이다. 제안하는 시스템을 위한 Hadoop 기반의 확장된 저장소 관리 모듈(Extended Storage Manager based on HDFS)과 정형 데이터 및 비정형 데이터에 대한 데이터 변환 모듈(Data Converter)로 구성된다. 확장 HDFS 저장소, 비정형 데이터 변환 모듈 등을 포함하며 비정형 데이터를 중심으로 데이터를 수집, 저장, 관리하며 분석에 활용한다. Fig. 3은 빅데이터 컨테이너의 구성을 보여준다.

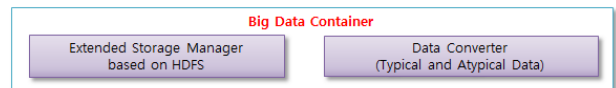


Fig 3. Big Data Container

분석 및 검증 컨테이너(Analysis & Verification Container)는 데이터를 분석 및 검증하고 시각화하는 기능을 가지며 기계학습 및 통계 기반의 데이터마이닝과 실험계획법을 기반으로 하는 데이터 분석 모듈과 분석 결과에 대한 시각화 모듈을 가진다. 데이터 분석 및 검증 모듈은 입력되는 데이터에 대하여 통계적인 분석, 실험계획법을 적용한 분석 및 검증 기능을 제공하는 분석(Analysis) 모듈, 입력되는 데이터를 기반으로 분석 결과에 대한 검증 기능을 제공하는 검증(Verification) 모듈, 입력되는 데이터를 기반으로 기계학습을 중심으로 하는 데이터마이닝 기술을 활용하여 데이터에 대한 분석 및 예측 기능을 수행하는 데이터마이닝 및 기계학습(Data Mining & Machine Learning) 모듈로 구성된다. 데이터 시각화(Data Visualization) 모듈은 분석, 검증 및 예측되는 데이터를 시각화하여 사용자에게 제공하는 기능을 수행한다.

회귀분석 및 계층적 군집 분석 등을 통한 데이터 분석과 신규 분석 모델의 개발을 병행한다. 분석된 모델을 대상으로 피드백을 수행하며, 피드백 결과에 의해 갱신된 정보는 개발된 신규 분석 모델의 검증을 위한 기준으로 활용한다.

데이터마이닝 기법, 실험분석법, 신경망과 기계학습 기반의 분석을 표준으로 하며, 시스템을 위한 신규 분석 모델 개발을 R통계분석 툴을 활용하여 자체 개발한 데이터 분석 API를 적용하여 데이터를 분석한다.

기존의 분석 기술과 기계학습 등을 통한 다양한 정량적 분석 및 예측 알고리즘 간의 교차검증과 피드백을 통하여 정성적, 정량적인 예측 결과 및 판단 근거를 모두 제시할 수 있는 검증 알고리즘을 적용한다. Fig. 4는 분석 및 검증 컨테이너의 구성을 보여준다.

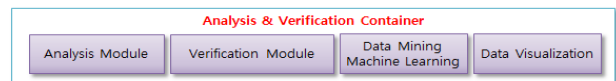


Fig 4. Analysis & Verification Container

산업 유형별 프로세스 컨테이너(Industry Process Container)는 다양한 산업 유형에 따른 공정 및 프로세스를 관리하는 컨테이너이다. 제조장비의 상태데이터, 제조 시점에서의 환경데이터, 작업지시정보 등 수집된 다양한 데이터에 대한 분석 및 검증 결과를 기반으로 해당 작업을 위한 표준 데이터 및 설정 값을 생성한다. Technical Standard Rule Manager, Data Level Manager, API Set Manager, Industry Process Manager로 구성한다.

기술 표준 규칙 관리자(Technical Standard Rule Manager)는 해당 산업 유형에 대한 기술 표준과 관련된 규칙을 관리한다. 데이터 등급 관리자(Data Level Manager)는 해당 산업 유형에 대한 각 데이터들의 난이도 및 보안 등급 등의 관리를 수행한다. API 셋 관리자(API Set Manager)는 산업

유형별 비즈니스 프로세스를 설계, 개발, 운영하기 위한 API 셋을 관리한다. 산업 유형별 프로세스 관리자(Industry Process Manager)는 산업 유형별로 정의된 비즈니스 프로세스를 구현하여 서비스를 제공한다. 비즈니스 프로세스의 구현은 API 셋을 이용하여 구현한다. Fig. 5는 산업 유형별 프로세스 컨테이너의 구성을 보여준다.

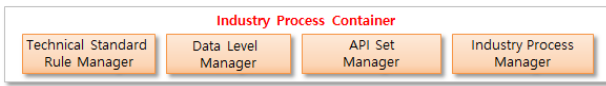


Fig 5. Industry Process Container

이러한 프레임워크를 적용함으로써 다음과 같은 장점을 기대할 수 있다.

- 프레임워크를 이용한 일관성 있는 시스템 개발 지원
- 추상화된 데이터 입력 인터페이스 지원
- 분산처리 기반의 데이터 처리 및 저장
- Raw 데이터(생산, 환경설계 작업데이터 등) 수집, 분석
- 통계, 데이터마이닝을 통한 상황 분석, 분류
- 정량적 검증 모듈을 이용한 수집 데이터 분석, 조정
- 일정 수준 이상의 품질 및 재사용성을 추구

3.4 데이터 분석/검증 및 시각화 기술

Fig. 6은 데이터 변환, 분석, 시각화 툴을 구성하는 요소들의 구조이다. Fig. 7은 수집된 데이터 소스로부터 데이터를 변환하고 분석한 후 시각화처리에 이르는 데이터의 흐름과 프로세스를 보여준다.



Fig 7. Structure of the Data Conversion, Analysis, and Visualization Tools

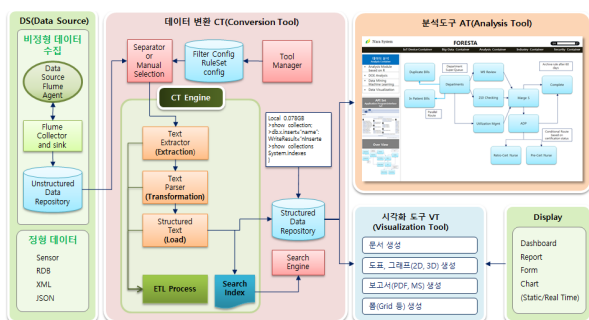


Fig 8. Data Flow and Process for the Data Conversion, Analysis, and Visualization Tools

데이터 변환 도구는 데이터를 Text Filter(Text Extractor)에서 읽고 원시 텍스트를 추출하고, Text Parser(Text Analyzer)에서 형태소 분석(정확성), 비정형 데이터 구분(엑셀 텍스트 길이 설정 유사), 전부 또는 특정 데이터(Trim 적용)를 분석한다. 분석 결과를 미리 정의된 다양한 형태(Json, Text, XML 등)로 저장한다. Batch Job(배치분석)를 통해 Batch 설정에 따른 주기적인 Conversion을 실행하고 Big Data일 경우 실시간 분석이 어려우므로 저장 후 준 실시간 분석을 수행하기 위해 배치로 실행하며, Batch Manager에서 관련 설정한다.

4. 제조공정 분석 및 시각화

데이터 분석을 공정 개선에 적용하기 위하여 현장에서 생산되는 데이터를 분석 및 검증한 후 작업자 또는 경영자에게 의사결정을 위한 데이터로 변환/제공하기 위하여, 통계 및 기계학습을 기반으로 데이터를 분석하고, 잘 구성된 시각화 기능을 적용하여 현장에서의 자원 및 정보의 흐름에 기반을 둔 실시간 관리 시스템이 요구되고 있다.

제조공정 정보 시각화란 사용자에게 더 효율적으로 정보를 전달하기 위하여 그래픽 요소를 활용하여 데이터가 정보로서 의미가 생성되도록 형상화하는 것이다[10, 15]. 본 실험에서는 다양한 원자재의 실제 배합 데이터 셋을 표본으로 하여 실험을 수행한다. 원자재 배합에 투입되는 정확한 재료명과 품목코드는 데이터 셋을 제공한 업체의 요청에 따라 대체코드로 표시하였다. 각 재료에 대한 대체코드는 4종류로 분류하고 각 분류에 따라 일련번호를 적용하여 대체코드를 생성하였다. Table 2는 대체코드를 나타낸다.

Table 2. Alternative Code

C1	KD120880014
C2	KD300000914
C3	KD120880012
C4	KD300000915

수집된 데이터는 분석 과정을 거쳐 제조공정에 피드백 됨으로써 그 효과를 확인할 수 있다. 분석 결과의 정확성 확보를 위하여, 수많은 사용자를 가지며 그 성능이 이미 검증된 오픈소스 통계 분석 툴인 R을 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, 분석 결과를 제조공정으로 피드백 함으로써 제조공정의 개선에 활용한다.

원자재 혼합 제조 시 적용되는 배합 비율표를 기준으로 정확한 조성을 위하여 보정한 실제 배합 데이터를 이용하여

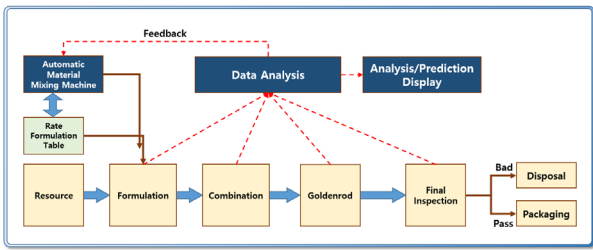


Fig 8. Data Analysis Flow for the Raw Materials

Table 3. Some of the Experiment Data for the Mixing Ratio of the Coatings

C1	C2	C3	C4	L	A	B	ΔE	Kg
2.814	20.167	46.9	101.304	27.78	-0.39	-1.51	0.918	454.93
2.814	23.067	46.9	101.3	27.8	-0.36	-1.14	0.782	454.93
3.907	23.1	47.12	101.3	27.45	-0.3	-1.15	0.445	454.93
4.554	32.595	75.18	164	27.5	-0.38	-1.16	0.506	733.194
4.78	32.6	75.2	164	27.49	-0.37	-1.16	0.495	733.194
1.068	7.654	17.8	38.5	27.69	-0.37	-1.17	0.685	171.948
1.068	7.654	18	38.5	27.51	-0.37	-1.17	0.517	171.948
1.168	7.714	18.04	38.5	27.5	-0.36	-1.16	0.502	171.948
2.109	15.165	35.26	76.248	27.53	-0.36	-1.09	0.510	342.41
2.1	15.2	35.3	79.3	27.49	-0.36	-1.09	0.472	342.41
3.04	21.758	50.6	124.5	27.51	-0.34	-1.61	0.789	506
3.04	24.758	50.6	124.5	27.53	-0.33	-1.28	0.578	506
3.04	26.76	50.6	124.5	27.53	-0.33	-1.01	0.495	506
3.8	27.27	63.4	156.1	27.49	-0.3	-1.54	0.719	634
3.8	30.27	63.4	156.1	27.5	-0.31	-1.25	0.535	634
3.8	33.27	63.4	156.1	27.51	-0.31	-1.08	0.482	634
3.8	27.262	63.42	137	27.41	-0.32	-2.52	1.585	614.98
3.8	37.762	63.42	137	27.45	-0.33	-1.92	1.027	614.98
3.8	47.762	63.42	137	27.47	-0.32	-1.34	0.563	614.98
3.82	52.9	63.5	137	27.42	-0.3	-1.12	0.406	614.98
0.48	3.42	7.96	17.2	27.75	-0.37	-1.32	0.794	77.117
0.48	3.42	10.96	17.2	27.45	-0.36	-1.31	0.534	77.117
0.74	4	11	17.2	27.46	-0.3	-1.15	0.454	77.117
2.14	15.275	35.5	87.33	27.81	-0.39	-2.48	1.690	355
5.14	25.275	45.5	87.33	27.58	-0.39	-1.39	0.689	355
5.14	28.275	50.5	87.33	27.46	-0.38	-1.22	0.496	355
5.14	28.275	50.6	90.4	27.45	-0.34	-1.21	0.475	355
7.055	50.575	117.5	253	27.66	-0.34	-1.53	0.832	1135.05
7.055	60.575	127.5	253	27.5	-0.34	-1.24	0.533	1135.05
7.055	66.575	127.5	253	27.51	-0.34	-1.17	0.512	1135.05
10.06	66.6	127.5	253	27.5	-0.29	-1.17	0.498	1135.05

실험을 수행하였다. 실험데이터는 제조현장에서 직접 측정된 실제 데이터를 적용하고 있으며, 데이터 셋을 제공한 업체의 현장상황에 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 환경정보를 적용하여 계속적으로 갱신되어야 한다. Fig. 8은 원자재의 배합 공정 데이터 분석 과정을 보여준다.

Table 3은 실험에 사용된 표본 데이터의 일부이다. Table 3에서 A4, B4, C1, C2, C4, C8, D3, D7, D14는 원자재 배합 시 투입되는 재료이며, 배합 결과물의 색상은 CCM을 통한 측정된 후 L, A, B, ΔE로 나타난다.

Table 4는 배합 결과물에 대한 색상의 방향성에 대한 기준표이며 기준에 따라 색상 값으로 배합에 따른 조색의 방향성을 확인할 수 있다. 실험데이터에 대하여 취합한 데이터는 회귀분석을 실시하여 적합성 여부를 확인한다.

Table 4. Direction of the Mixing Color

	L	A	B
PLUS	White	Red	Yellow
MINUS	Black	Green	Blue

R-square 값이 0.6보다 크면 그 모형이 신뢰할만한 모형이라고 판단된다. 각 값 L a b deltaE에 대해 회귀분석을 실시하고 예측모형을 설계한다. 다음으로 실제 데이터 같은 Input value를 넣어 예측을 실시하여 실제 데이터 결과 값과 비교한다.

Fig. 9는 실험데이터 분석 그래프이다. 전체적인 범위는 예측 값과 실제 값이 유사한 범위에 있으나 어긋난 데이터들을 확인할 수 있다.

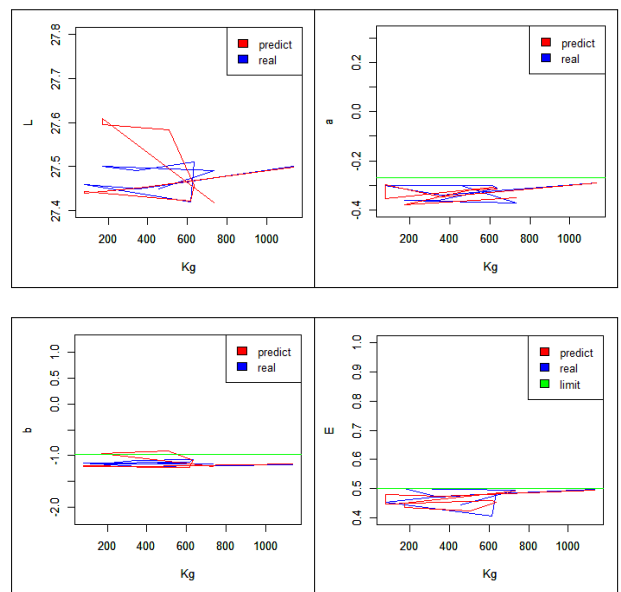


Fig 9. Graph of Prediction of the Experimental Data

Fig. 10은 실험 데이터의 분석을 통하여 관계식을 도출하고, 관계식에 따라 보정된 데이터를 이용하여 다시 예측을 수행한 결과이다. Fig. 9와 비교할 때 오차 범위가 매우 줄어든 것을 확인할 수 있다. Fig. 10에서 파란 점이 실제 데이터의 값이고 빨간 점이 예측된 데이터의 값이다.

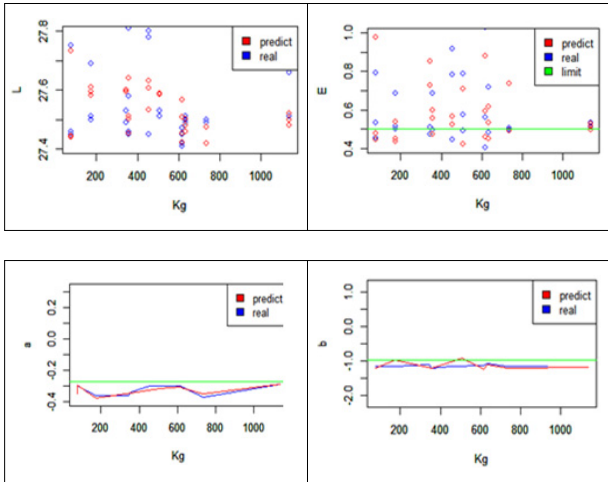


Fig 10. Graph of Predict of the modified experimental data

위 그래프에는 적합한 색상을 찾아가는 과정이 담겨있다. 적합도 E가 0.5 이하이면 그 색상은 적합한 색상이라고 판단되어 0.5 이하의 데이터만 비교해 다음과 같은 그래프를 도출하였다.

Lab 그래프는 초록색 선에 가까울수록 적합하고 E 그래프는 초록색 선보다 밑에 있으면 적합하다고 볼 수 있다.

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 정확도가 증가할 것이다.

일반적인 분석 결과 데이터와 분석 그래프의 경우 실제 작업 시에 보고 빠르게 이해하기 힘든 경우가 많으므로 직관적으로 최적화의 범위를 보여줄 필요가 있다. 본 논문에서 제안하는 프레임워크를 이용하여 웹 기반의 데이터 분석 서비스를 개발하였으며, 다음과 같은 환경에서 운영한다.

- Server : Tomcat 7.0
- 데이터 분석 모델 및 API 구현 : Java
- 외부 웹 인터페이스 : PHP, Java Script
- 기본 통계알고리즘 : GNU-R 원격접속

Fig. 11은 프레임워크를 이용하여 개발한 데이터 분석 서비스의 화면이다. 입력 데이터를 선택하면 입력된 세 가지 원료의 데이터를 이용하여 분석을 수행한다.

Fig. 12는 분석된 데이터를 각 항목별 기준을 통하여 최적화 지점이 어디인지 보여주는 화면이다.

분석이 완료된 후 본 논문에서 제안하는 제조공정 분석서

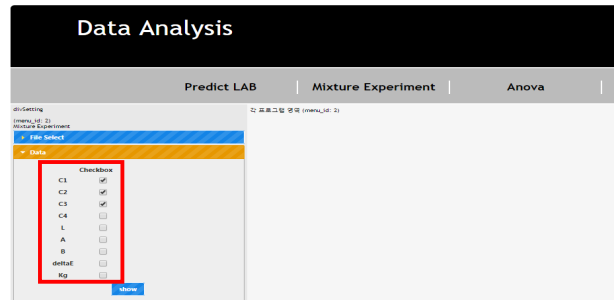


Fig 11. An Example of Mixture Experiment Function in Developed Data Analysis Service System

비스의 선택 메뉴에서 다양한 방법을 사용해 정밀한 분석을 할 수 있다.

화면의 최상단에서 각 원료와 오차율(E)의 범위를 지정할 수 있다. 회귀분석의 정도를 달리하여 데이터의 수에 따라 유동적인 분석이 가능하다.

Linear 옵션은 1차 회귀분석, Quadratic 옵션은 2차 회귀 분석, Cubic은 3차 회귀분석으로 차수가 커질수록 필요 데이터 양이 많아지지만, 신뢰도가 높은 결과를 볼 수 있다. 또한 범례의 개수를 선택할 수 있고 값의 정밀도를 높여 더욱 정밀한 값을 볼 수 있다.

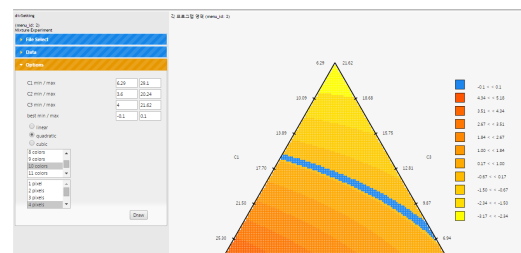


Fig 12. An Example of Mixture Experiment Function in Developed Data Analysis Service System

Fig. 13은 분석된 데이터를 각 항목별 기준을 통하여 최적화 지점이 어디인지 다양한 특성 기준에 따라 3D로 보여주는 화면이다.

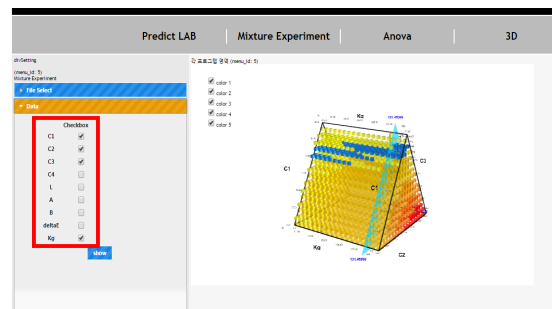


Fig 13. An Example of Mixture Experiment 3D Function in Developed Data Analysis Service System

Fig. 14는 분석 결과를 이용하여 입력데이터에 따른 예측

결과와 실제 결과를 비교할 수 있는 예측비교 그래프이며, Fig. 15는 분석된 데이터의 값을 직접 보여주는 화면이다.

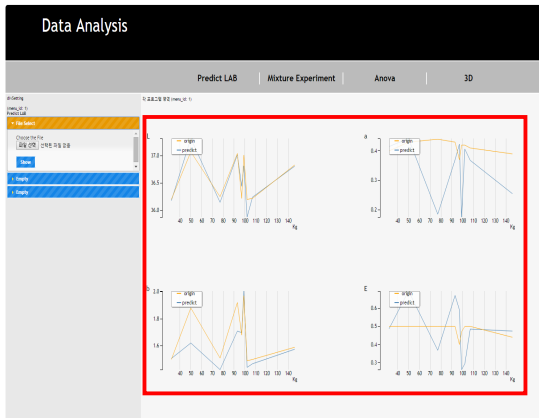


Fig 14. An Example of Prediction Function

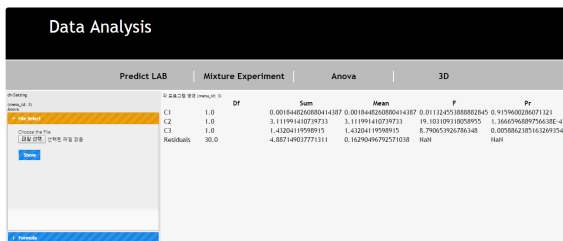


Fig 15. The Result Data of Analysis

실험을 통하여 실제 생산 현장에서 사용하는 기준 데이터, 인적 오류, 제품의 특성 값 및 기준 값 등에 발생한 오차에 대하여 작업자가 수정한 보정 결과 데이터를 실험계획법을 기반으로 분석하여 기준 데이터를 계산하였다.

데이터 분석을 통하여 원자재 배합에서 가장 중요한 인자를 검색하고 기준 값을 보정하여 최적화 값을 도출하였으며, 도출한 데이터 최적화 모형에서 배합공정에서의 원료 코드별 지시량과 불출량, 자재의 기본 색의 영향, 자재가 가진 특성의 중요도에 대하여 실험계획법의 하나인 혼합물 실험 계획법을 통해 데이터의 방향성과 배합의 연관성, 결과물과 목표 간의 오차 감소 방향 등을 분석하였다.

5. 결론

본 논문에서는 실험계획법과 데이터 분석 기반의 원자재 혼합물의 배합 비율 최적화를 위한 정보 시각화를 제안하였다. 또한, 실험과 데이터 분석을 통하여 원자재 배합에서 가장 중요한 인자를 검색하고 기준 값을 보정하여 최적화 값을 도출하였으며, 확보된 기준 데이터를 작업자에게 쉽게 제공하고, 제조공정에 적용하기 위하여 정보 시각화 기술을 적용한 데이터 분석 서비스를 개발하였다. 작업 데이터의 분석을 통한 배합 비율의 최적화와 이해하기 쉬운 분석 내용의 제공은 원자재 배합의 정확도 향상과 작업시간 단축을

가능하게 해주고, 건당 처리시간의 감소로 인한 생산 납품 시간 단축, 불량률 감소 등에 따른 원가 절감에 기여할 수 있다. 뿐만 아니라, 다양한 모델링에 대한 제조공정에서의 표준 데이터를 획득할 수 있다.

향후에는 제안된 최적화 모형의 데이터 처리 아키텍처를 기반으로 개발하는 시스템에 본 연구를 통하여 축적된 데이터를 적용하고, 도출된 결과를 이용하여 고객 요구에 맞는 색차 값 추출 및 표준 값과 제품의 특성 값을 비교, 분석 및 검증한 후 최적화를 위한 모형에 따른 데이터를 제시하는 다양한 상황에 맞는 표준 TA를 설계하고자 한다. 또한 수집된 자료를 데이터 검증 분석 방법에 적용하여 표준 아키텍처의 제시 및 객관적 검증을 수행할 수 있는 기술에 대한 연구에 중점을 두고자 한다.

References

- [1] Eizo Kinoshita and Takao Ohya, Strategic decision-making technique AHP, Cheongram, 2012.
- [2] Sung-Hyun Park and Jong Wook Kim, "Modern Design of Experiments using MINITAB," Minyoungsa, Seoul, 2011.
- [3] Kae Gil Lee, "Statistical analysis using the MATLAB," Ajin, Seoul, 2013.
- [4] Hyeyoung Park and Kwanyong Lee, "Pattern Recognition and Machine Learning: From the ground up to take advantage of," Ehan Media, Goyang, 2011.
- [5] Michael J. Crawley, "The R Book," 2nd ed., Acorn Publishing, 2014.
- [6] Jung, D. B. and Won, T. Y, "Time series data and analysis of simplification I," Hannarae Publishing Group, Seoul, 2001.
- [7] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining: Concepts Techniques," Acorn Publishingm, 2015.
- [8] Brett Lantz, "Machine Learning with R," Acorn Publishing, 2014.
- [9] Richard O. Duba, Peter E. Hart, and David G. Stork, "Pattern Classification," 2nd Ed., ITC, 2006.
- [10] Byungkeun Oh and Sung Joong Kang, "Textbook of information design," Ahn Graphics Publishers, 2008.
- [11] Chunghyun Yu and Seonghak Hong, "R Visualization," Insight, 2015.
- [12] Functional Inorganic Coatings [Internet], http://mirian.kisti.re.kr/publication/view.jsp?record_no=155&cont_cd=KI, 2007.
- [13] Heewon Jeon, "Big Data Analysis with R: R Distributed Programming," Microsoftware, pp.186-193, Sep., 2011.
- [14] 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.
- [15] H. S. Shin, J. M. Lim, and J. S. Park, "Information Visualization and Information Presentation for Visually Impaired People," *Electronics and Telecommunications Trends*, Vol.28, No.1, pp.81-91, 2013.



김재천

e-mail : k971110@hanmail.net
2006년 부경대학교 컴퓨터공학과(석사)
2015년 부경대학교 컴퓨터공학과(박사)
현 재 ㈜나라시스템 대표이사
관심분야: 부실채권 관리, 데이터마이닝,
프레임워크, 포렌식



노성여

e-mail : sungnsn@naver.com
1998년 부경대학교 전산정보(석사)
2014년 부경대학교 기술경영(박사수료)
현 재 동명대학교 항만물류시스템학과
교수
관심분야: 항만물류시스템, 데이터 분석,
프레임워크



진선아

e-mail : mystarta@naver.com
2010년 부경대학교 컴퓨터멀티미디어공학
(학사)
현 재 ㈜나라시스템 기업부설연구소
선임연구원
관심분야: 부실채권 관리, 데이터마이닝,
프레임워크, 포렌식



이현동

e-mail : win4calss@hanmail.net
2007년 부경대학교 컴퓨터공학과(석사)
2012년 부경대학교 컴퓨터공학과(박사)
현 재 ㈜나라시스템 기업부설연구소
연구소장
관심분야: 컴퓨터응용보안, 상황인식 컴퓨팅



박영희

e-mail : cafe0107@hanmail.net
2006년 부경대학교 컴퓨터공학과(석사)
2015년 부경대학교 컴퓨터공학과(박사)
현 재 한국폴리텍대학 정보통신시스템과
조교수
관심분야: 데이터마이닝, 부실채권 관리