

R기반 데이터 분석 프레임워크를 이용한 코팅제 배합 분석 기술

노성여[†], 김민정^{**}, 김영진^{***}

An Analysis Techniques for Coatings Mixing using the R Data Analysis Framework

Seong Yeo Noh[†], Minjung Kim^{**}, Young-Jin Kim^{***}

ABSTRACT

Coating is a type of paint. It protects a product forming a film layer on the product and assigns various properties to the product. Coating is one of the fields which is being studied actively in the polymer industry. Importance of coating in various industries is more increased. However, mixing process has been performing in dependence on operator's experience. In this paper, we found the relationship between each data from coating formulation process. We propose a framework to analyze the coating formulation process as well. It can improve the coating formulation process. In particular, the suggested framework may reduce degradation and loss costs due to absence of standard data which is accurate formulation criteria. Also it suggests responses to errors which can be occurred in the future through the analysis of the error data generated in mixing step.

Key words: Data Analysis, Data Verification, Framework, Coatings Mixing Process, Optimization

1. 서 론

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 [2,3]. 이에 배합 공정 현장에서는 지능화되고 자동화된 처리 기술의 필요성이 대두되고 있다. 최근 다양한 산업의 융합이 주류를 이루고 있다. 코팅제의 경

우는 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. 코팅제는 다양한 산업분야에서 활용되는 중요한 재료중 하나이며, 완성된 제품의 성능과 디자인적인 요소의 향상을 위하여 정확한 색상과 품질이 유지되어야 한다. 본 논문에서는 R 분석으로 코팅제 배합 공정을 분석하기 위하여 프레임워크를 활용하고 분석한 결과는 보다 정량적인 작업 기준 데이터를 확보하여 작업 현장에 제공함으로써 코팅제 배합 공정을 개선시킬 수

* Corresponding Author: Young Jin Kim, Address: (608-711) 179, Sinseonno, Nam-Gu, Busan, Korea, TEL: +82-51-629-6486, E-mail: youngk@pknu.ac.kr
Receipt date: Mar. 4, 2015, Revision date: Apr. 8, 2015
Approval date: Apr. 17, 2015

[†] Dept. of Interdisciplinary Program of Management of technology, Pukyong National University
(E-mail: sungnsn@naver.com)

^{**} Nara System Co., Ltd.
(E-mail: k971110@hanmail.net)

^{***} Dept. of Systems Management and Eng., Pukyong National University

* This work (Grants No. C0219239) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2014.

있다. 특히 정확한 배합 기준이 되는 표준 데이터의 부재로 인한 품질 저하와 원가 손실을 감소시키고, 배합 공정에서 발생한 오차 데이터에 대하여 표준 보정 관계식을 도출함으로써 차후 발생 가능한 오차에 대한 대응 방안을 제시하여 최적의 코팅제 배합 기준 데이터를 확보할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구 및 배경 기술을 소개하고 3장에서는 프레임워크를 활용하여 분석된 데이터를 기반으로 한 코팅제 배합 공정 데이터에 대한 분석결과를 제시한다. 마지막으로 4장에서는 결론과 향후 연구에 대하여 논의한다.

2. 관련 연구 및 배경 기술

2.1 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 사용자에 의하여 충분히 검증된 분석 프로그램이다. In-Memory 컴퓨팅 기술을 통한 빠른 처리속도와 분석 및 시각화를 위한 강력한 그래픽 함수를 지원한다. 본 논문에서는 R을 기반으로 다양한 데이터의 분석과 해당 데이터의 신뢰성을 검증하는 API 셋을 제안한다. 제안하는 API 셋은 프레임워크의 분석 및 검증 컨테이너의 구성요소 중 하나이다[4].

2.2 데이터 마이닝(Data Mining)

많은 데이터 가운데 숨겨져 있는 유용한 상관관계

를 발견하여, 미래에 실행 가능한 정보를 추출해 내고 의사 결정에 이용하는 과정을 말한다. 데이터 마이닝에는 인공지능의 한 분야이자 자동적 학습 기법을 설계 및 구현하는 기계학습, 데이터에서 유용한 패턴을 찾아내는 패턴인식, 통계학, 신경망 컴퓨팅 등의 기술이 포함된다. 본 논문에서는 코팅제 배합 분석기술을 기준으로, 경영자 및 현장의 작업자에게 배합공정의 최적화를 지원하기 위하여 각 데이터에 대한 관계성을 통계적 기법 및 데이터 마이닝 기법을 이용하여 분석하고 그 결과를 제공한다.

2.3 코팅제 배합공정

코팅제는 수지, 안료(조색제), 용제, 첨가제로 구성된다. 수지는 코팅제의 기본 성분으로 도막의 기능 중에서 화학적, 물리적 기능을 좌우한다. 안료는 용제에 용해되지 않고 분산되어 있는 무기 화합물을 말하며, 색상 특징을 부여하고, 도막의 기계적 성능을 개선하거나 다양한 효과를 부여한다[1]. 첨가제는 코팅제의 제조에서 시작하여 코팅제가 건조되고 내구성을 가질 때까지 다양한 기능을 가질 수 있도록 해 주는 보조적인 기능을 가진다[1]. 단일 안료만으로 구성된 코팅제를 혼합하여 원하는 색상으로 조합하는 작업을 조색이라 한다. 조색방법은 조색자의 기술 숙련도가 품질 형성에 중요한 역할을 하는 육안조색법과 컴퓨터 매칭시스템을 사용하는 CCM (Computer Color Matching)[5] 활용 조색법으로 구분된다. Fig. 1은 자동 코팅제 배합 공정의 데이터 흐름도이다.

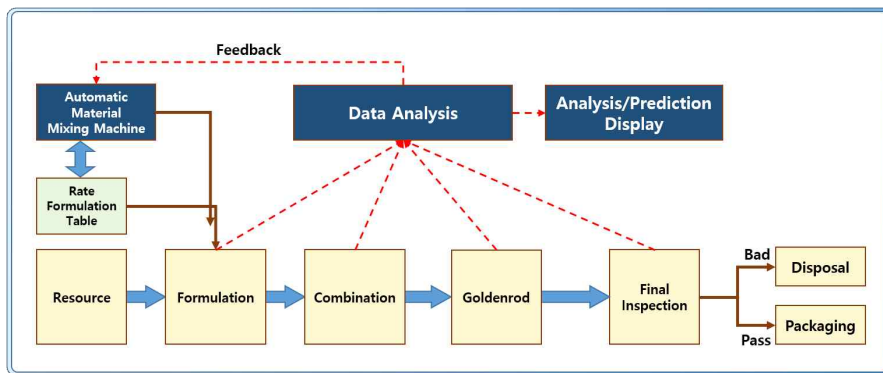


Fig. 1. Data flow of the automatic coatings mixing process.

3. 코팅제 배합공정 데이터 분석

3.1 데이터 분석 및 검증 프레임워크

프레임워크의 사용은 소프트웨어 재사용을 통한 비용절감 및 개발기간 단축, 데이터 분석에 관한 표준 API의 지원에 따른 일관성 유지 등의 장점을 가진다. 뿐만 아니라, 표준 분석 프로세스를 확립하고, 다양한 도메인에서의 요구사항을 반영 및 각 도메인에 대한 기술구조 확립이 가능하다. Fig. 2는 코팅제 배합공정 데이터 분석에 사용한 FORESTA 프레임워크의 아키텍처이다[5].

수집된 데이터는 분석 및 검증 컨테이너를 통해서 데이터 마이닝, 신경망 및 기계학습 기반으로 분석된다. Fig. 3은 FORESTA 프레임워크의 분석 및 검증 컨테이너의 아키텍처이다. 데이터 소스(DS)는 비정형(JSON 파일 외), 정형 데이터(NC 외)로 구성되며, 데이터 변환(CT)에서 Text Filter에서 데이터를 읽고 원시 텍스트 추출, Text Parser에서 형태소 분석을 통해 정확성, 데이터 길이, 특성을 구분하여 분리한다. Rule Set은 분석된 정보를 정의된 형태로 저장하며, 결과 검색을 위한 Index를 생성한다. Tool Manager(Batch Job, Real-time Job, Interactive Job)는 주기적인 Conversion을 실행하고 실시간 분

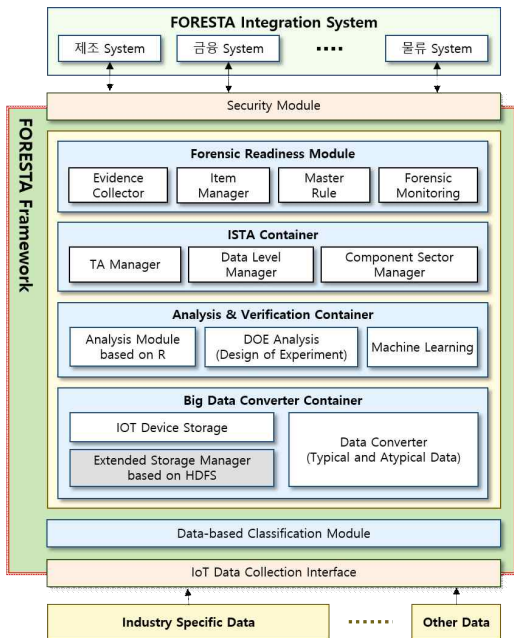


Fig. 2. FORESTA framework[5].

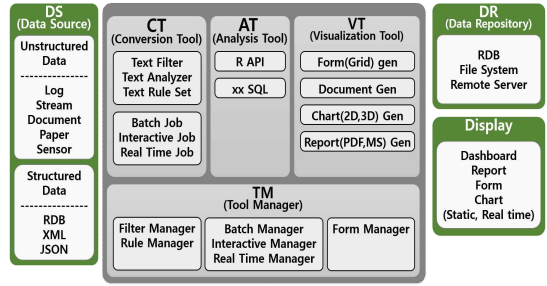


Fig. 3. Analysis & verification container of FORESTA framework[5].

석이 어려운 데이터는 저장 후 배치로 실행해서 검증을 하며, 사용자의 요청에 의한 실행 작업은 Rule Set 설정 후 작업을 진행한다.

개발 시스템을 위한 R분석 모델 기반은 데이터마이닝과 시스템을 위한 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 TA 설계를 한다. 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 TA를 설계한다.

3.2 코팅제 배합 공정 데이터 분석

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 재료의 기본구성(용제, 수지, 조색제, 첨가제)의 4종류로 분류하고 각 분류에 따라 일련번호를 적용하여 대체코드를 생성하였다. Table 1은 재료의 구분

Table 1. The configuration of alternative code for the material mixing of the coatings

Material classification	Code	Number
Solvent	A	1 ~ 4
Resin	B	1 ~ 4
Concentrates	C	1 ~ 9
Additives	D	1 ~ 14

에 따른 코드 생성 규칙이다.

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. 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.

Table 2는 1차 실험에 사용된 데이터 중 코팅제 #01에 대한 표준 데이터의 일부이다. Table 2에서 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 3의 기준에 따라 색상의 수치 값으로 배합에 따른 조색의 방향성을 확인할 수 있다. Fig. 4는 실험 데이터에 대한 R

Table 2. First partial experimental data

	1	2	3	4	5
A4	7.298	7.298	7.298	7.298	7.298
B4	51.976	51.976	51.976	51.976	51.976
C1	42.72	47.32	52.32	57.32	60.72
C2	2.136	2.136	2.136	2.136	2.136
C4	3.56	3.56	3.56	3.56	3.76
C8	58.74	66.34	66.34	73.34	76.34
D3	1.234	2.234	3.934	4.634	4.834
D7	1.78	1.78	1.78	1.78	1.78
D14	3.56	3.56	3.56	3.56	3.56
Total	173.004	186.204	192.904	205.604	212.404
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
ΔE	4.26	3.22	1.43	0.69	0.22

Table 3. Direction of the mixing color

	L	A	B
PLUS	White	Red	Yellow
MINUS	Black	Green	Blue

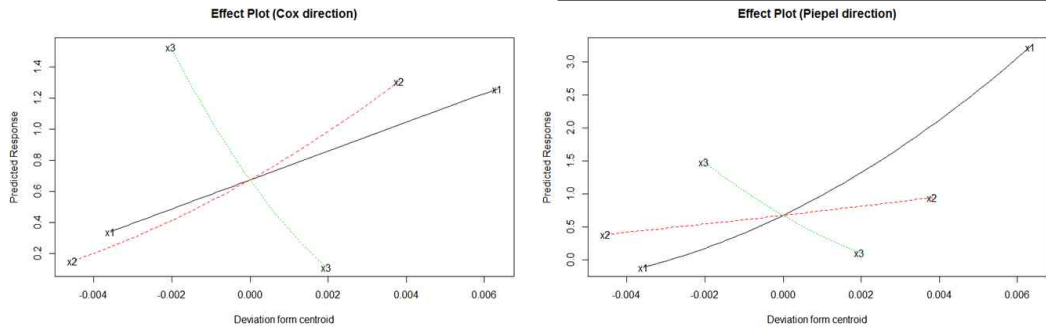


Fig 4. Cox reaction trajectory and Piepel reaction trajectory by R.

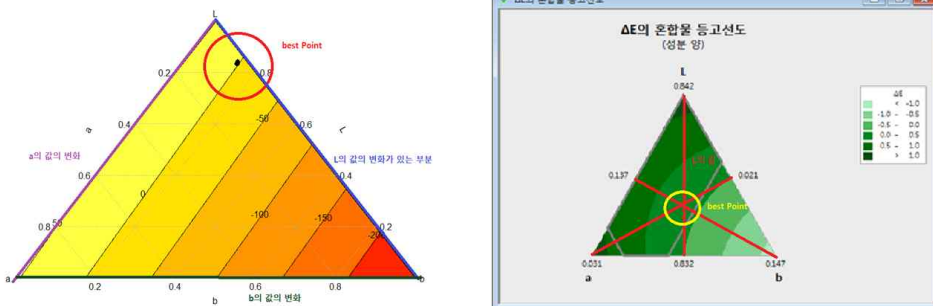


Fig. 5. The contour matrix for the mixing material by R(left) and MiniTab(right).

Table 4. Second partial experimental data

C1	C2	C3	C4	L	A	B	ΔE	Kg
2.814	20.167	46.9	101.304	27.78	-0.39	-1.51	0.918096	454.93
2.814	23.067	46.9	101.3	27.8	-0.36	-1.14	0.781857	454.93
3.907	23.1	47.12	101.3	27.45	-0.3	-1.15	0.44486	454.93
4.554	32.595	75.18	164	27.5	-0.38	-1.16	0.506063	733.194
4.78	32.6	75.2	164	27.49	-0.37	-1.16	0.494874	733.194
1.068	7.654	17.8	38.5	27.69	-0.37	-1.17	0.684544	171.948
1.068	7.654	18	38.5	27.51	-0.37	-1.17	0.51672	171.948
1.168	7.714	18.04	38.5	27.5	-0.36	-1.16	0.502096	171.948
2.109	15.165	35.26	76.248	27.53	-0.36	-1.09	0.510196	342.41
2.1	15.2	35.3	79.3	27.49	-0.36	-1.09	0.471911	342.41
3.04	21.758	50.6	124.5	27.51	-0.34	-1.61	0.789113	506
3.04	24.758	50.6	124.5	27.53	-0.33	-1.28	0.577668	506
3.04	26.76	50.6	124.5	27.53	-0.33	-1.01	0.494571	506
3.8	27.27	63.4	156.1	27.49	-0.3	-1.54	0.719027	634
3.8	30.27	63.4	156.1	27.5	-0.31	-1.25	0.534883	634
3.8	33.27	63.4	156.1	27.51	-0.31	-1.08	0.482183	634
3.8	27.262	63.42	137	27.41	-0.32	-2.52	1.584614	614.98
3.8	37.762	63.42	137	27.45	-0.33	-1.92	1.027278	614.98
3.8	47.762	63.42	137	27.47	-0.32	-1.34	0.563028	614.98
3.82	52.9	63.5	137	27.42	-0.3	-1.12	0.406079	614.98
0.48	3.42	7.96	17.2	27.75	-0.37	-1.32	0.793536	77.117
0.48	3.42	10.96	17.2	27.45	-0.36	-1.31	0.533948	77.117
0.74	4	11	17.2	27.46	-0.3	-1.15	0.454093	77.117

기반의 Cox반응 궤적도와 Pieple 반응 궤적도이다. Fig. 5는 실험 데이터에 대한 R과 미니탭의 혼합등고 선도이다.

Table 4는 2차 실험에 사용된 데이터의 일부이다. Fig. 6은 실험 데이터를 R을 이용하여 회귀분석을 수행한 결과화면이다.

회귀분석 결과에서 R-square 값이 0.6보다 크면 그 모형이 신뢰할만한 모형이라고 판단된다. 각 값 L, a, b, ΔE에 대해 회귀분석을 실시하고 예측모형을 설계한다. 다음으로 실제 데이터 같은 Input value를 넣어 예측을 실시하여 실제 데이터 결과 값과 비교한다. Fig. 7은 실험데이터에 대한 분석 그래프이다. 파란 점이 실제 데이터의 값이고 빨간 점이 예측된 데이터의 값이다. 위 그래프는 적합한 색상을 찾아가는 과정이 담겨있다. 적합도 E가 0.5이하이면 그 색상은 적합한 색상이라고 판단되어 0.5이하의 데이터만 비교해 다음과 같은 그래프가 도출되었다. Lab 그래프는 초록색 선에 가까울수록 적합하고 E 그래프는 초록색 선 보다 밑에 있으면 적합하다.

예측된 결과 값과 실제의 결과 값을 비교해보았을

때 차이가 근소하며 때로는 예측된 값이 더 좋은 결과를 나타낸다. 또한 더 많은 데이터를 이용하여 회

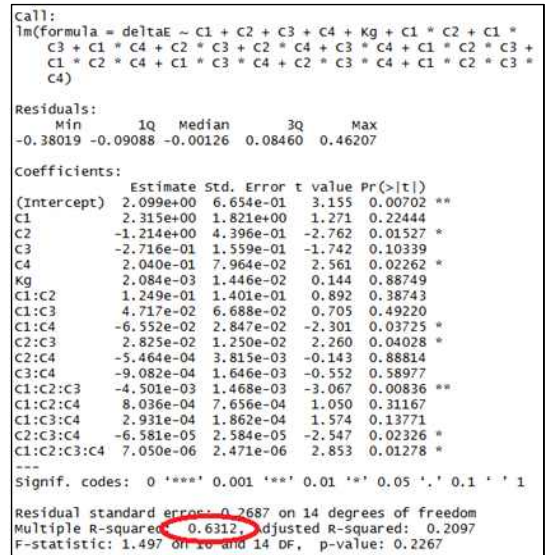


Fig 6. Regression analysis of the experimental data using R.

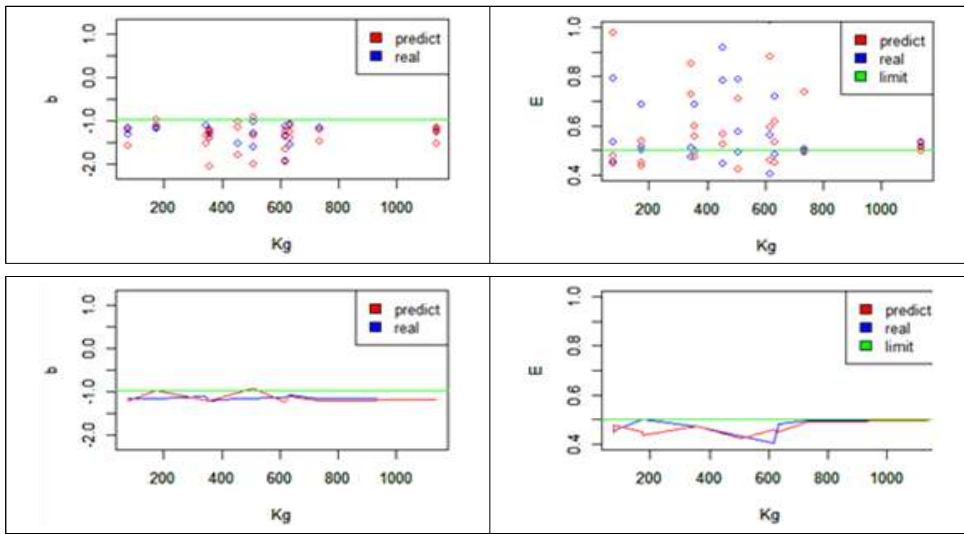


Fig 7. Graph of analysis of the experimental data.

귀분석을 실시하면 예측에 대한 정확도 또한 증가할 것이다.

4. 결 론

본 논문에서는 R을 기반으로 하는 데이터 분석 프레임워크를 이용한 코팅제 배합 분석 기술을 제안하고 그 실험결과를 제시하였다. 1차 실험을 통하여 코팅제 배합 시 정배합 샘플값에서 ΔE 의 오차범위 안의 Best Point 범위, 조색제 각각에 대한 교호작용 및 점밀성은 전혀 무관함을 실험에서 얻을 수 있었다. 2차 실험을 통해 Best Point 범위 내 분석결과에 따른 예측값, 실제값, 비교결과, 정확성을 통한 표준 한도 색상값과 거의 일치한 값을 획득하였다. 본 논문에서 도출한 코팅제 배합비율 데이터 최적화 모형에서 원료 코드별 지시량과 불출량, 도료의 기본색의 영향, 조색 시 조색제와 수지의 중요도에 대하여 조색 데이터의 방향성과 배합의 연관성, 결과물과 목표 간의 오차 감소 방향 등을 분석하였다. 본 연구에서의 실험과 분석을 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.

향후에는 데이터 처리 아키텍처를 기반으로 시스템에 축적된 데이터를 적용하고, 도출된 결과를 이용하여 고객 요구에 맞는 색차 값 추출 및 표준 값과

제품의 특성 값을 비교, 분석 및 검증하는 표준 구조 및 시각화를 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.

REFERENCE

[1] Functional Inorganic Coatings, http://mirian.kisti.re.kr/publication/view.jsp?record_no=155&cont_cd=KI (accessed Nov., 5, 2014).

[2] Analyst G. Johnson Interviewed by S. Bushell in *Computer World*, 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] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2008.

[5] J.C. Kim, *Design of Framework for the Data Analysis and Validation Based on the Forensic Readiness by Industry Type*, Doctor's Thesis of Pukyong National University, 2015.

[6] M. Koo, J. Lee, and H. Cha, *CCM System*

- Practical Technology Research for Laboratory Dyeing Machine*, Korean Institute of Industrial Technology, 1993.
- [7] J.C. Stanley, "The Influence of Fisher's 'The Design of Experiments' on Educational Research Thirty Years Later," *American Educational Research Journal*, Vol. 3, No. 3, pp. 223-229, 1966.
- [8] J.F. Box, "R.A. Fisher and the Design of Experiments," *The American Statistician*, Vol. 34, No. 1, pp. 1-7, 1980.
- [9] H. Bae, S.S. Kim, D.W. Choi, S.T. Lee, and Y.J. Kim, "Application of Data Mining for Coagulant Dosage of Water Treatment Plants Corresponding to Input Conditions," *Journal of Korean Institute of Intelligent Systems*, Vol. 15, No. 1, pp. 1-5, 2005.
- [10] S.R. Kim and M.M. Kang, "Today and the Future of Big Data Analytics Technology," *Proceeding of Communications of the Korean Institute of Information Scientists and Engineer*, pp. 8-17, 2014.
- [11] W.S. Han, "Social and Technological Big Data Analytics Technology," *Proceeding of Communications of the Korean Institute of Information Scientists and Engineer*, pp. 3-5, 2014.
- [12] I.K. Oh, Y.M. Kim, and S.W. Cha, "A Study on the Application Methods of Big Data in CMF Design Process of Smart Phone," *Journal of Digital Design*, Vol. 14, No. 4, pp. 781-790, 2014.
- [13] Y.Y. Park and H.W. Byun, "Efficient Development and Quality Evaluation of Online Library System Using Framework," *Journal of Korea Multimedia Society*, Vol. 13, No. 5, pp. 627-640, 2010.
- [14] J.H. Lee and S.M. Hwang "Value-Based Enterprise Architecture Framework: VBEAF," *Journal of the Korea Industrial Information Systems Research*, Vol. 19, No. 6, pp. 77-79, 2014.
- [15] K.W. Lim and Y.W. Kim, "Performance Analysis of Data Processing Framework based on Hadoop," *Korean Society for Internet Information Summer Announcement Conference*, Vol. 13, No. 1, pp. 225-226, 2012.
- [16] Y.H. Kang and M.G. Hong, "Sensory Data Processing by Using Hadoop Framework," *The Journal of Korean Institute of Information Technology*, Vol. 11, No. 2, pp. 169-174, 2013.
- [17] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2008.
- [18] H.W. Jeon, "Big Data Analysis with R: R Distributed Programming," *Microsoftware*, pp. 186-193, 2011.
- [19] K.G. Lee, *Statistical Analysis using the MATLAB*, Ajin, Seoul, 2013.
- [20] H.Y. Park and K.Y. Lee, *Pattern Recognition and Machine Learning: From the Ground up to Take Advantage of*, Ehan Media, Goyang, 2011.
- [21] D.B. Jung and T.Y. Won, *Time Series Data and Analysis of Simplification I*, Hannarae Publishing Group, Seoul, 2001.
- [22] J.H. Lee, "Data Bigbang, Trends of Big Data" *Journal of Communications & Radio Spectrum*, No. 47, p. 45, 2012.
- [23] Y.J. Choi and S.S. Park, "Interplay of Text Mining and Data Mining for Classifying Web Contents," *Korean Journal of Cognitive Science*, Vol. 13, No. 3, pp. 33-46, 2002.
- [24] J.J. Lee, *A Study on the Realtime Ascii Art Works Using Processing Language*, Master's Thesis of Chung-Ang University of Multimedia & Film, 2007.



노 성 여

현재 동명대학교 항만물류시스템
학과 교수
2014년 부경대학교 (기술경영)대
학원 박사 수료
1998년 부경대학교 (전산정보)대
학원 석사 졸업

관심분야: 항만물류 시스템, 데이터 분석, 프레임워크



김 영 진

현재 부경대학교 시스템경영공학
부 교수
2000년 Ph.D in Industrial
Engineering, Clemson
University
1992년 한국과학기술원(KAIST)
경영과학 공학석사

1987년 한국과학기술원(KAIST) 경영과학 이학사
관심분야: 응용통계, 품질공학, 최적화 기법의 응용, 경
제성 분석



김 민 정

현재 (주)나라시스템
2013년 동아대학교 (전자공학과)
대학원 석사 졸업
관심분야: 상황인식