

Design of Data Fusion and Data Processing Model According to Industrial Types

Min-Seung Jeong[†] · Seon-A Jin^{††} · Woo-Hyun Cho^{†††}

ABSTRACT

In industrial site in various fields it will be generated in combination with large amounts of data have a correlation. It is able to collect a variety of data in types of industry process, but they are unable to integrate each other's association between each process. For the data of the existing industry, the set values of the molding condition table are input by the operator as an arbitrary value When a problem occurs in the work process. In this paper, design the fusion and analysis processing model of data collected for each industrial type, Prediction Case(Automobile Connect), a through for corporate earnings improvement and process manufacturing industries such as master data through standard molding condition table and the production history file comparison collected during the manufacturing process and reduced failure rate with a new molding condition table digitized by arbitrary value for worker, a new pattern analysis and reinterpreted for various malfunction factors and exceptions, increased productivity, process improvement, the cost savings. It can be designed in a variety of data analysis and model validation. In addition, to secure manufacturing process of objectivity, consistency and optimization by standard set values analyzed and verified and may be optimized to support the industry type, fits optimization(standard setting) techniques through various pattern types.

Keywords : Data Fusion, Data Mining, Data Processing Technology, Manufacturing Process, Device

산업유형별 데이터융합과 데이터처리 모델의 설계

정 민 승[†] · 진 선 아^{††} · 조 우 현^{†††}

요 약

다양한 분야의 산업 현장에서는 복합적으로 대용량의 데이터가 상호 연관성을 가지고 발생한다. 산업유형별 공정에서 다양한 데이터들을 수집할 수 있으나, 각 프로세스 사이에서 서로 연관성 있게 통합하지 못하고 있다. 기존 산업유형별 데이터는 성형조건표 설정치 값과 작업공정에서 문제가 발생 했을 경우 작업자가 임의 값을 입력하였다. 본 논문에서는 각 산업유형별로 수집되는 데이터의 융합 및 분석처리 모델의 설계를 하고, 예측 사례(자동차 커넥트)를 통해서 표준 성형 조건표를 통한 마스터 데이터와 제조공정 과정에서 수집된 생산이력파일을 비교분석하여 다양한 불량요인과 예외사항에 대한 패턴분석과 재해석, 작업자에 대한 임의 값을 수치화를 통해 새로운 성형 조건표를 통한 불량률 감소, 생산성 증가, 공정 개선, 원가 절감 등의 기업 수익 향상과 제조 산업의 공정에 맞는 다양한 데이터분석과 검증 모델을 설계할 수 있다. 또한, 분석 검증된 표준 설정치에 의한 제조 공정의 최적화, 일관성, 객관성을 확보할 수 있고 다양한 패턴유형을 통한 산업유형별에 맞는 최적화(표준 설정치) 기술을 지원할 수 있다.

키워드 : 데이터 융합, 데이터 마이닝, 데이터 처리 설계, 제조 공정, 디바이스

1. 서 론

제조 현장에서 발생하는 대량의 데이터는 활용하기에 따라 매우 중요한 역할을 수행할 수 있다. 특히 제조 공정의 개선

등을 위해서는 데이터에 대한 적절한 분석 기술과 분석 결과를 현장 및 경영자에게 빠르고 효율적으로 전달할 필요가 있다.

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 대기업과 그 1차 협력사와는 달리, 국내 제조업체의 대부분을 차지하는 2, 3차 협력사의 경우 업체의 규모 및 장비의 영세성 때문에 작업자의 수작업과 육안 검사에 의해 품질 검

※ 이 논문은 부경대학교 자율창의학술연구비(2015년)에 의하여 연구되었음.

† 준 회 원 : 부경대학교 컴퓨터공학과 석사과정

†† 정 회 원 : ㈜나라시스템 연구원

††† 종신회원 : 부경대학교 컴퓨터공학과 교수

Manuscript Received : October 10, 2016

Accepted : November 3, 2016

* Corresponding Author : Woo-Hyun Cho(whcho@pknu.ac.kr)

사를 진행하는 경우가 대부분이다. 또한, 제조환경 및 공정 데이터에 대한 객관적인 분석이 지원되지 않아 숙련자의 공백이 발생할 경우 비숙련자에 의한 인적 오류의 발생률도 증가하게 된다[1].

기술구조(Technical Architecture)란 세부적인 요구 사항을 충족시키는 적합한 체계(Conformance System)를 보장하기 위하여 체계를 구성하는 부품 또는 요소 간의 배열(Arrangement), 상호 작용 또는 상호 독립성을 지배하는 규칙의 최초 집합을 명시한 것으로서, 표준, 규정 및 지침 등으로 구체화된다[2].

현재 다양한 실시간 자동 관측 분야에서 USN(Ubiquitous Sensor Network)이 사용되고 있다. USN을 바탕으로 한 실시간 모니터링 시스템은 관측 장소에 구성된 센서 네트워크로부터 데이터를 수신 받아 별도의 서버 시스템에 데이터베이스를 구성하고, TCP 기반의 비표준적인 별도 프로토콜을 이용하는 모니터링 프로그램이 데이터베이스의 자료를 수신하여 표출하는 형태를 갖춘다[3].

본 논문에서는 제조 공정에서 발생한 생산 데이터, 휴먼 에러, 예외사항, 불량 데이터에 대한 분석을 통하여 최적화를 도출하고 차후 발생할 수 있는 불량률, 오차율 등의 예외사항을 예측하고 공정 개선을 지원할 수 있는 다양한 분석과 검증 방법을 통해 최적화 할 수 있는 기술 구조를 제시하고, 최적화된 기술구조를 피드백, 실증을 통해 제조 공정을 지원할 수 있는 산업유형별 데이터융합과 데이터처리 모델의 설계를 연구하고자 한다.

본 논문의 구성은 다음과 같다. 2장 관련 연구에서는 데이터 수집 디바이스, 데이터분석 및 데이터 셋을 살펴보고, 3장에서는 데이터융합과 데이터 모델 설계, 전처리 설계, 데이터 분석 API 도구 설계, 산업유형별 데이터처리 과정, 표준 TA를 활용한 불량 예측 사례를 살펴보고, 4장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

2.1 제조 공정 데이터 수집 디바이스

ARTIK 10은 삼성전자가 개발한 애플리케이션 프로세서(AP), 통신칩, 센서 등으로 구성되는 개방형 IoT 플랫폼으로 구성된 개방형 사물인터넷(IoT)이다[4]. ARTIK 10을 Fig. 1에 나타내었다.

oneM2M 공통 플랫폼은 레이어 아키텍처 스타일로서 표현하면 M2M 어플리케이션 레이어와 하단의 M2M 네트워크 서비스 레이어 사이에 위치하고 있으며, 네트워크 서비스 레이어가 제공하는 네트워크 서비스를 이용하면서 M2M 기반 다양한 어플리케이션들에게 공통 기능을 제공하기 위한 공통 서비스 레이어(CSE)로 표현될 수 있다[5].

oneM2M 플랫폼 아키텍처는 기능 엔티티 모델로 표현이 가능하며 센서, 액추에이터, 디바이스, 게이트웨이가 위치하는 필드 도메인 영역과 서비스 제공자의 서버가 위치하는 인프라스트럭처 도메인 영역으로 구분이 가능하다. 또한, 어

플리케이션 레이어에 위치하는 Application Entity(AE), 공통 서비스 플랫폼 레이어에 위치하는 Common Service Entity(CSE), 그리고 네트워크 서비스 레이어의 Network Service Entity(NSE)로 분리하여 기능 모델을 정의하고 있다. oneM2M 공통 플랫폼 기능 모델을 Fig. 2에 나타내었다.

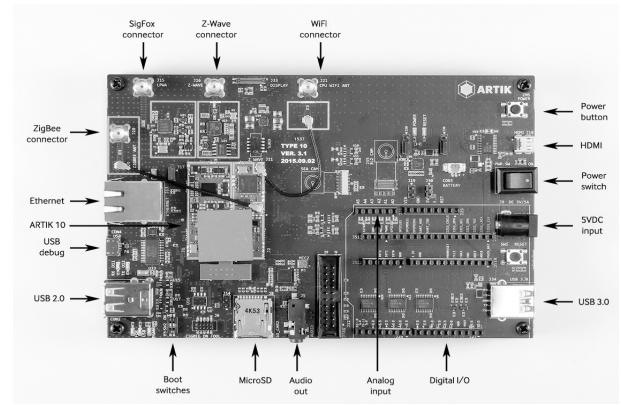


Fig. 1. SAMSUNG IoT 'ARTIK 10'

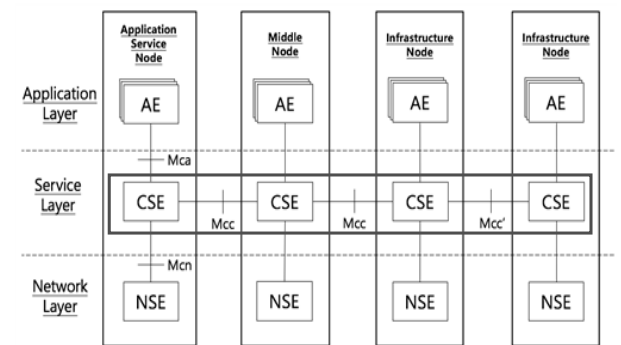


Fig. 2. oneM2M Common Platform Functional Model

2.2 데이터웨어하우스의 구축을 위한 측정지표 설계

데이터웨어하우스는 경영자의 의사 결정을 지원하는 주제지향적(Subject-oriented), 통합성(Integrated), 시계열성(Timevariant) 그리고 비휘발성(Nonvolatile) 데이터의 집합체이다[6].

- 주제지향적: 데이터를 주제별로 구성함으로써 최종 사용자(end user)와 전산에 약한 분석자라도 이해하기 쉬운 형태로 유지한다.
- 통합성: 데이터가 데이터웨어하우스에 들어갈 때는 일관적인 형태(데이터의 일관된 이름 명시, 일관된 변수 측정, 일관된 코드화 구조 등)로 변환되어 데이터의 통합성이 유지된다.
- 시계열성: 데이터웨어하우스의 데이터는 일정 기간 동안 정확성을 유지한다.
- 비휘발성: 데이터웨어하우스에 일단 데이터가 적재되면 일괄처리(batch) 작업에 의한 갱신 이외에는 「Insert」나 「Delete」 등의 변경이 수행되지 않는다[7].

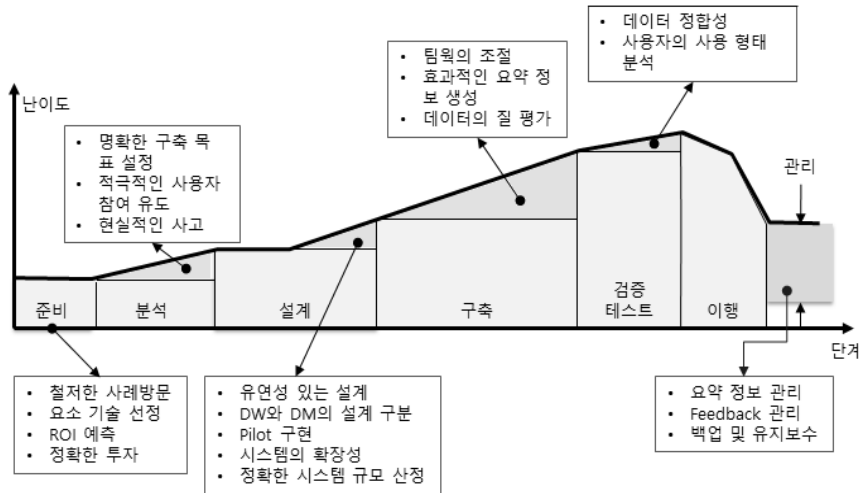


Fig. 3. Measurement Index for Building Data Warehouse

데이터웨어하우스는 관리 하드웨어, 관리 소프트웨어, 추출·변환·정렬 도구, 데이터베이스 마케팅 시스템, 메타 데이터(meta data), 최종 사용자 접근 및 활용 도구 등으로 구성된다. Fig. 3에 데이터웨어하우스 구축을 위한 측정 지표를 나타내었고, Table 1은 측정항목을 나타내었다[8].

Table 1. Measurement Index of Data Warehouse

division	Metrics	Explanation	
productivity	S	Size of software	
	E	The airstrike into the project (Man/Month)	
	D	D1	Project plan days
		D2	Project Delayed Days
Process quality	R	R1	Number of defects removed during development
		R2	Number of defects removed after delivery to customer
		R	Number of defects removed (R=R1+R2)
	A	A1	Number of requirements change
		A2	Requirements Implementation Count
A		Number of requirements	
Data quality	F	F1	Number of defects detected during development
		F2	Number of defects discovered after delivery to customer
		F	Total number of defects in the project (F=F1+F2)
		Data	Quality of the data itself
Data quality	Data management	Data Management Policy	
	Data structure	Structure between data	

2.3 데이터 분석 처리 및 데이터 셋

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 고용의 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.

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[1].

지금까지 로지스틱 회귀분석을 이용한 연구로는 단변량 및 다변량 분석기법을 이용하여 금융, 마케팅, 고객 만족도 조사 등 여러 분야에 대한 연구들이 있다[9].

선형 회귀분석은 선형성을 가지는 독립변수와 종속변수의 관계를 설명하거나 두 변수의 관계성을 이용하여 데이터의 변화를 예측하는 통계방법이다. 즉 회귀분석에서 독립변수에 따라 종속변수의 값이 일정한 패턴으로 변해 가는데, 이러한 변수간의 관계를 나타내는 회귀선이 직선에 가깝게 나타나는 경우를 선형회귀분석이라고 한다[1, 6]. 독립변수가 하나인 경우를 단순회귀분석, 여러 개인 경우를 다중회귀분석이라고 한다[1].

- 단순회귀분석: 가장 단순한 분석 모형으로 두 개의 변수만으로 이뤄져 있고, 오차 항이 있는 선형관계로 모형을 나타낸다.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- 다중회귀분석: 두 개 이상의 독립변수를 사용해 하나의 종속변수의 변화를 나타낸다.

$$y_i = \beta_0 + \beta_1 u_i + \beta_2 v_i + \beta_3 w_i + \epsilon_i$$

2.4 산업유형별 자동차 커넥터 제조 공정

산업유형별 자동차 커넥터 제조 공정은 원재료 수입 검사, 원재료 보관, 분쇄 및 압출, 혼합, 원재료 건조, 원재료 투입, 금형 정비 및 체결, 사출 성형-조건 관리, 검사 관리, 포장, 창고 적재 및 출하 단계로 이루어진다.

Fig. 4에 자동차 커넥터 제조 공정 단계를 나타내었다. 이러한 제조 공정 단계에서 수집된 성형조건표 설정값, 불량 이력, 환경정보, 작업 데이터 등의 데이터 집합을 데이터 셋이라고 한다. Fig. 5에 특별특성표와 불량이력 샘플을 나타내었다.



Fig. 4. Automotive Connector Manufacture Process

Quality Characteristics Control Plan																																							
Serial No.	Item No.	Item Name	Item Description	Material	M/C TYPE	M/C #1	M/C #2	Unit	Lot Weight	Lot Size																													
001	001	001	001	001	001	001	001	001	001	001																													
<table border="1"> <tr> <th>Process</th> <th>Main Control Items</th> <th>Control Type</th> <th>Control Method</th> <th>Control Point</th> </tr> <tr> <td>Material</td> <td>001</td> <td>001</td> <td>001</td> <td>001</td> </tr> <tr> <td>Operation</td> <td>001</td> <td>001</td> <td>001</td> <td>001</td> </tr> <tr> <td>QC</td> <td>001</td> <td>001</td> <td>001</td> <td>001</td> </tr> <tr> <td>Finishing</td> <td>001</td> <td>001</td> <td>001</td> <td>001</td> </tr> <tr> <td>Other</td> <td>001</td> <td>001</td> <td>001</td> <td>001</td> </tr> </table>										Process	Main Control Items	Control Type	Control Method	Control Point	Material	001	001	001	001	Operation	001	001	001	001	QC	001	001	001	001	Finishing	001	001	001	001	Other	001	001	001	001
Process	Main Control Items	Control Type	Control Method	Control Point																																			
Material	001	001	001	001																																			
Operation	001	001	001	001																																			
QC	001	001	001	001																																			
Finishing	001	001	001	001																																			
Other	001	001	001	001																																			
<table border="1"> <tr> <th>Inspection Flow</th> <th>1. DEL. 삽입부 SAGE 검사</th> <th>2. DEL. 삽입부 측정</th> <th>3. DEL. 1차 LOCKING</th> <th>4. T.M. 삽입</th> <th>5. DEL. 2차 LOCKING</th> <th>6. 삽입부 확인</th> </tr> <tr> <th>Critical Defect</th> <td>1. PUSH-LOCK부</td> <td>2. CORE BROKEN</td> <td>3. CORE BROKEN</td> <td>4. PUSH-IN HOLE</td> <td>5. PUSH-IN HOLE</td> <td>6. DEL. OIL</td> </tr> </table>										Inspection Flow	1. DEL. 삽입부 SAGE 검사	2. DEL. 삽입부 측정	3. DEL. 1차 LOCKING	4. T.M. 삽입	5. DEL. 2차 LOCKING	6. 삽입부 확인	Critical Defect	1. PUSH-LOCK부	2. CORE BROKEN	3. CORE BROKEN	4. PUSH-IN HOLE	5. PUSH-IN HOLE	6. DEL. OIL																
Inspection Flow	1. DEL. 삽입부 SAGE 검사	2. DEL. 삽입부 측정	3. DEL. 1차 LOCKING	4. T.M. 삽입	5. DEL. 2차 LOCKING	6. 삽입부 확인																																	
Critical Defect	1. PUSH-LOCK부	2. CORE BROKEN	3. CORE BROKEN	4. PUSH-IN HOLE	5. PUSH-IN HOLE	6. DEL. OIL																																	

Fig. 5. Special Characteristics and Failure History

3. 데이터융합과 데이터처리 모델의 설계

3.1 데이터 수집용 디바이스 설계 및 개발

본 논문에서 제안하는 데이터 수집 디바이스는 제조 공정에서 취득 가능한 제품에 대한 특성 값, 초기 값 등의 생산 데이터와 센싱 데이터를 수집한다.

다양한 데이터를 주기적으로 수집, 저장할 수 있으며, 다양한 제조장비 및 제조군에 대한 표준 값과 오차율을 시각화 저작도구에 지원할 수 있고, 서버에 분산 저장, 분석/검증을 통하여 갱신이 가능할 수 있게 설계되었다. (Fig. 6)에 데이터 수집용 디바이스 설계를 나타내었다.

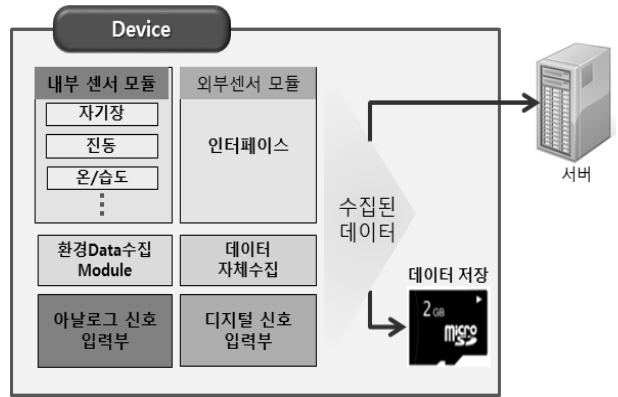


Fig. 6. The Device Structure and Data Flow

데이터 수집용 디바이스는 온도, 습도, 전압, 전류 측정을 위한 센서 모듈과 아날로그 데이터의 출력을 수행하는 모듈로 구현되어 있다. 또한 생산 데이터와 이력파일을 저장할 수 있는 SD카드를 탑재하고 있다. Fig. 7에 데이터 수집 디바이스 실물을 나타내었고, Fig. 8에 펌웨어와 Main.c 소스 파일을 나타내었다.

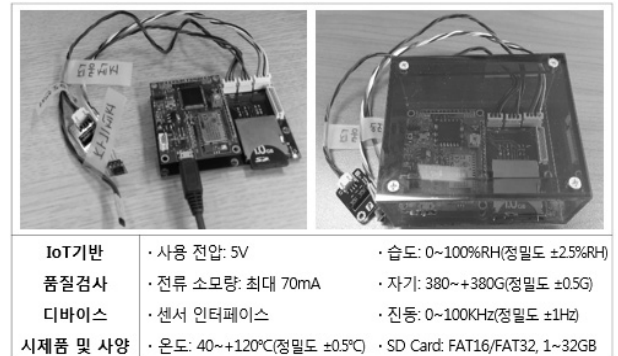


Fig. 7. The Actual Data Acquisition Device

```

//-----//
// main.c : PSOC Main Program
// Target : CY8C58LP
// Clock : 24MHz Xtal
// Revision : 1.0
//-----//

#include <device.h>
#include <FS.h>
#include <string.h>
#include <Global.h>
#include "stdio.h"

#define MAX_SAMPLE 10
#define LCD_NUM_COLUMNS 16u
#define BUFFER_LEN 64u
#define Mag_Offset 70.
#define ASCII_LF 0x0A
#define ASCII_CR 0x0D

char sdFile_Tmp[12] = "LOGFILE.TMP";
char sdFile[12] = "LOGFILE.TXT";
char sdVolName[10];
UI6 step = 1u;
FS_FILE * pFile;
    
```

Fig. 8. Firmware Main.c Source File

3.2 데이터 전처리 설계

설정된 기준 마스터 데이터와 생산 데이터(이력)를 읽어 서 Text로 추출한 후 Filtering 과정을 거쳐 정제된 데이터

를 생성한다. 이때 다양한 데이터(정형, 비정형 외)를 빅데이터 분석기를 통해 Text 파일을 생성한다.

Text화 작업 후 1차 분석과정을 거친 Text는 측정항목 등을 기록한 인덱스가 생성된 파일로 저장된다. 제품 생산에서 관리 될 항목과 불량예측, 휴먼에러, 제품 생산에 미칠 수 있는 요인 등에 대한 분석을 데이터 분석/검증 모듈에서 처리한다. Fig. 9에 전처리 처리과정을 나타낸다.

Text Filter(Text Extractor)에서 데이터의 원시 텍스트 추출, Text Parser(Text Analyzer)에서 데이터 분석 부분, Rule Set(Text Combination)에서 분석된 정보를 정의된 형태로 저장하는 부분, Batch Job(배치분석), Interactive Job(사용자 요청에 의한 분석), Real-time Job(실시간 분석)의 전처리 단계를 거친다. Fig. 10에 데이터 전처리 프로세스 처리 과정을 나타낸다.

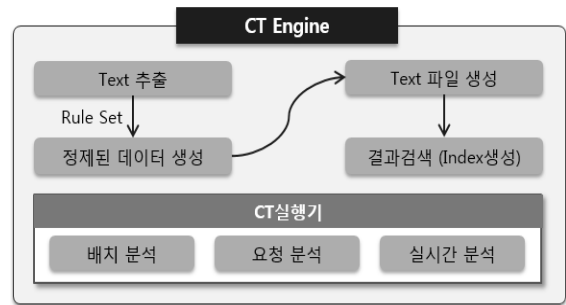


Fig. 9. Data Preprocessing

3.3 데이터 분석 API 도구 설계

제조 공정에서 수집된 데이터는 분석 과정을 거쳐 제조 공정에 피드백 됨으로써 그 효과를 확인할 수 있다. 분석 결과의 정확성 확보와 실험을 통하여 도출된 보정방정식과 계수

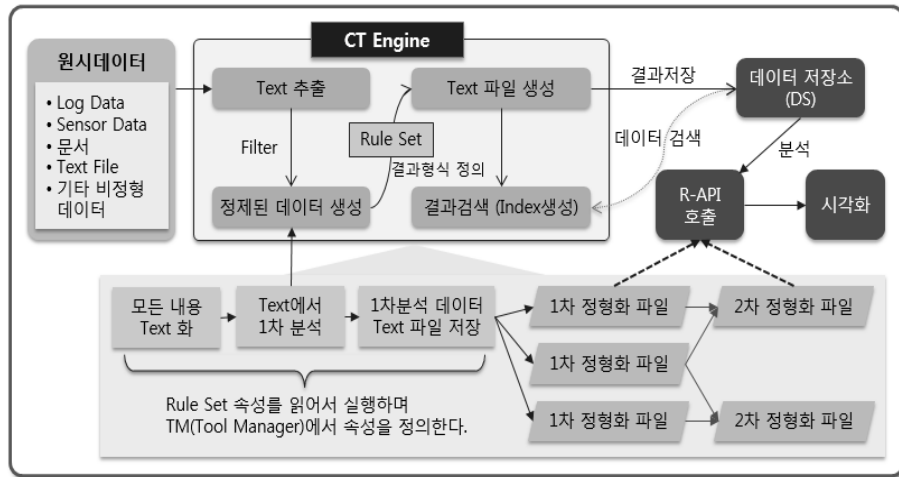


Fig. 10. Process and Structure of Data Preprocessing

```

//텍스트 추출
TextExtractor txtExt = new TextExtractor();

//변수 선언
String exText = "";
String anText = "";

// DS에서 텍스트 추출
exText = TextExtractor.TextExtract(filePath);

// 텍스트 분석(형태소분석 이면 형태소분석 처리)
if("Y".equals(fkAnalysisYN))
    anText = TextAnalyzer.TextAnalyzer(exText, fSeparator);

// 텍스트 조합
String cbText = TextCombination.TextCombine(anText, fSeparator, fCombination);

// 결과 mongoDB에 저장
TextLoader txtLoader = new TextLoader(mongoTemplate);

//JSON 저장위한 Bean에 결과값 세팅
ConvDoc doc = new ConvDoc(fileName, filePath, softDate.setCurDatetime(), fConvType, cbText);

//결과 출력
jsonList.add(doc);
    
```

Fig. 11. Text Analysis Algorithm

를 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 설계하였다.

데이터의 분석/검증이 종료된 후 생성된 표준 정보를 데이터 수집용 디바이스에 저장하고 해당 데이터를 활용하여 제조 과정에 최적화된 환경을 유지할 수 있게끔 설계하였다.

데이터 분석 API 도구 설계는 데이터 소스, 데이터 분석, 산업유형별 기술구조, 디스플레이, 데이터 저장소 등의 컴포넌트로 구성된다.

데이터 소스 컴포넌트는 다양한 형식의 데이터를 분석과 검증을 할 수 있는 형식으로 변환, 새로운 변수(측정항목)를 생성한다. 데이터 분석/검증 컴포넌트는 분석 및 검증을 수행할 수 있도록 R 통계분석 툴을 기반으로 데이터마이닝, 기계학습 등의 API 모듈로 구성되어 있다.

분석/검증을 마친 데이터 셋과 관계식을 기반으로 산업유형별 컴포넌트는 다양한 프로세스에 적용하기 위한 최적화 지침서, 규정, 조건, 설정치를 제공하며, 최적화된 데이터를 시각화 저작도구(모니터링)을 통해 작업자 및 관리자에게 제공된다. 또한, 마스터 데이터, 이력파일, 인덱스 처리된 가공 데이터를 각각의 특성에 맞는 데이터 저장소에 저장된다.

기존의 분석 방법, 데이터마이닝, 기계학습 등을 통한 다양한 분석방법을 선정하고, 기존 패턴분석(모델)과 재해석 및 신규 분석 모델을 교차검증과 피드백을 통하여 정성적, 정량적인 예측 결과 및 판단근거를 모두 제시할 수 있는 검증 알고리즘을 설계하고 적용한다. Fig. 12에 데이터 분석

Table 2. The List of Data Analysis API

Clustering	kmeans	K-means clustering
	hclust	Hierarchical clustering
	cutree	Truncate tree into groups of data
Machine Learning	predict	prediction
	lm	Aligning linear models
	princomp	Principal component analysis
	lda	Linear discriminant analysis
Design of Experiment	predict.lm	Classify multivariate observations by linear discrimination
	interaction.plot	Graphic interactions of linear models
	crvtave	Make pole design
	SCD	Creation of mixture design based on simplex center
	Effplot	Generate mixture effect plot
	ModelPlot	Contour plots of functions and functions provided in a simplex mixture space

및 검증 과정을 나타내었고, Table 2에 다양한 데이터 분석 API 항목 예시를 나타내었다.

교차검증과 피드백을 통하여 예측된 데이터 분석 결과값을 다양한 그래프 형태의 시각화를 통해 제조업체의 작업자 및 경영자에게 보다 효율적으로 전달할 수 있고, 시각화 저작도구(모니터링)를 활용하여 설정 값에 대한 예측과 변동률을 보여 줄 수 있다. 기계학습 분석 결과 그래프, Mixture Experiment_3D(단면 절단) 등을 Fig. 13(A~C)에 데이터 시각화 예시를 나타내었다.

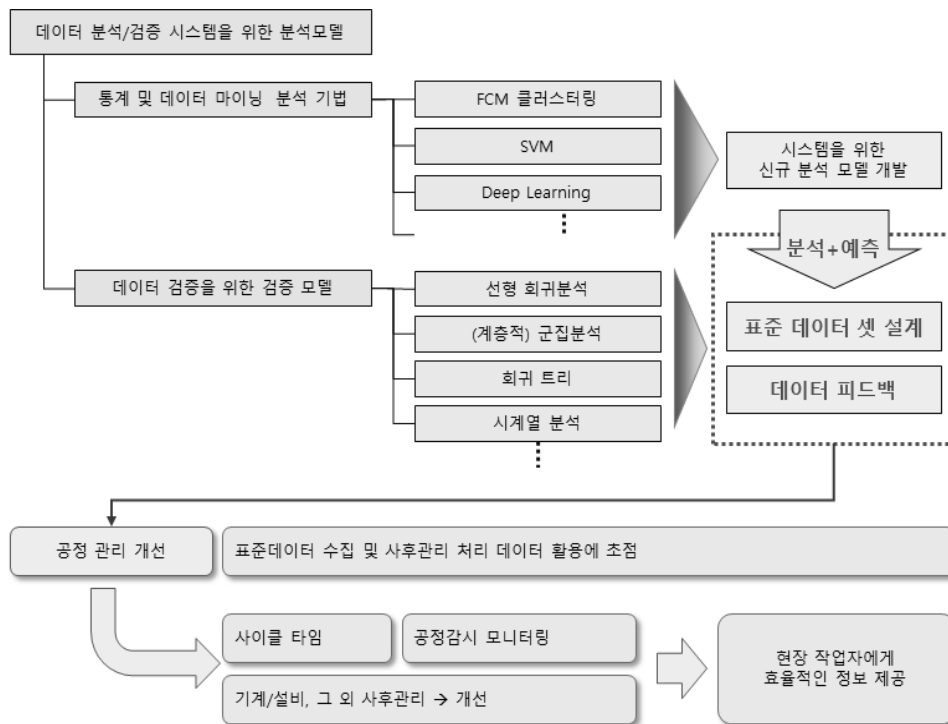


Fig. 12. Data Analysis and Verification Process

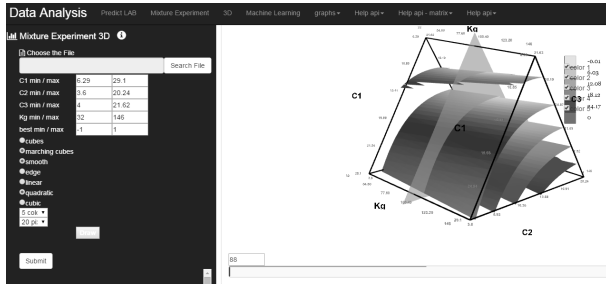


Fig. 13A. Visualization Graph Illustrates

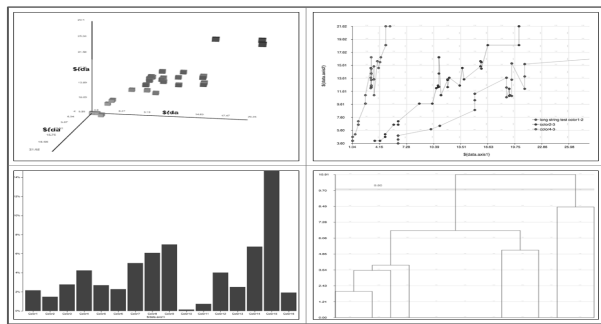


Fig. 13B. Visualization Graph Illustrates

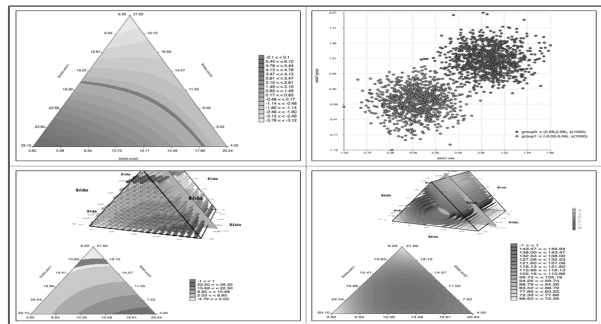


Fig. 13C. Visualization Graph Illustrates

```

KmeansClustering <- function(DataSet, ClusterNum) {
  // sample를 뽑아 첫 seed 설정
  select.no <- sample(nrow(DataSet), ClusterNum, replace=FALSE)
  ClusterCenter <- matrix(0, ClusterNum, ncol(DataSet))
  for(i in 1:ClusterNum) {
    ClusterCenter[i, 1:ncol(DataSet)] <- DataSet[select.no[i], 1:ncol(DataSet)]
  }

  while(TRUE) {
    LastClusterCenter <- ClusterCenter
    // 가장 가까운 중심의 그룹에 배속
    for(i in 1:nrow(DataSet)) {
      for(j in 1:ClusterNum) {
        // 편차계산
        Deviation[j, i] <- sqrt(sum((DataSet[i, 1:ncol(DataSet)] - ClusterCenter[j, 1:ncol(DataSet)])^2))
      }
      Minimum <- min(Deviation)
      for(j in 1:ClusterNum) {
        // 가장 가까운 집단에 배속
        if(Deviation[j, i] == Minimum) {
          Result[i, j] <- j
          break
        }
      }
    }

    // 중심 중심의 변화 확인
    if(sum((LastClusterCenter - ClusterCenter)^2) == 0) {
      return(Result)
    }
    n <- n+1
  }
  return(Result)
}
    
```

Fig. 14. Visualization Graph Algorithm

3.4 산업유형별 데이터처리 과정

데이터처리 과정의 최종 목표는 제조 공정에서 수집한 제조 장비 및 품목에 대한 표준 데이터(환경정보, 원자재 분석, 작업 지시 정보, 제조 품목의 정보 및 오차/불량률 분석 정도 등)를 저장하고 추출하여 관리한다.

설정된 표준 데이터 셋 기준으로 분류 후 서버로 전송하고, 분석과 검증된 데이터 결과 값은 지속적으로 갱신 반복 과정을 거친다. 또한, 피드백 및 실증을 통하여, 공정 불량 및 품질 예측의 정확도를 향상시킨다. Fig. 15에 산업유형별 데이터 처리과정의 흐름도를 나타내었다.

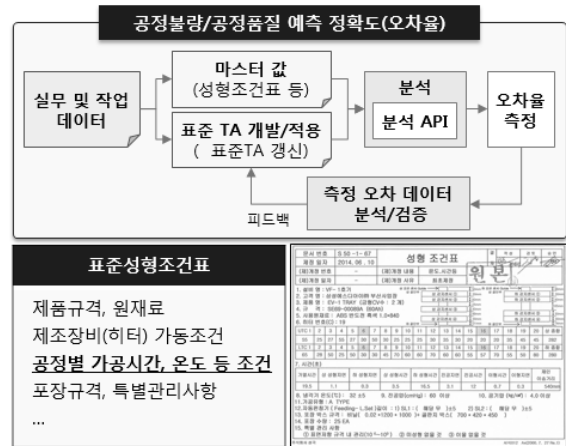


Fig. 15. Industry Type Data Processing Flow

실험데이터 및 마스터(샘플) 데이터 비교방법은 제조회사 A사의 특성을 따르며, 사출성형은 약 3~4초마다 제품이 생산, 데이터 크기 약 4M 정도의 데이터가 발생한다. 사출성형기 60대, 한 달 이상 정도의 데이터를 사출성형 기계업체에서 보관하고 있으며, 전체 용량은 테라바이트 크기이다. 자료 추출 및 보안 관련으로 인한 사출성형 샘플 시 자료를 획득했으며 제품군은 10개 이상, 샘플데이터는 약 3개월 기간 자료를 수집하였다.

- 마스터(샘플) 데이터: 자동차 커넥터 제조회사 A사의 표준 성형조건 설정치 항목(사출압력, 사출속도, 보압력, 보압시간, 계량위치, 보압전환, 냉각시간, 배압, 회전수, 온조기 온도, Logic 온도 등)의 표준값을 사출성형기에 입력된 값을 의미한다.
- 실험데이터: 표준 성형조건 설정치에 설정된 값을 모의 실험 후 제품에 영향을 줄 수 있는 관리항목을 중점으로 성형기에서 추출하였으며, 배치 작업을 통한 자료 수집을 하였다.
- 실험결과: 마스터 데이터의 설정치를 통한 생산이력파일을 분석하여 불량률과 예외사항 등에 대한 비교분석을 하였다. 분석결과 제조업체에서 요구하는 마스터 데이터의 일치성은 생산기계마다 각각 달리 나왔으며, 표준성형 설정치 값에 있어 관리항목 중 냉각속도에 가장 민감하였다. 작업자의 공정에서의 입력값에 대한 신뢰성은

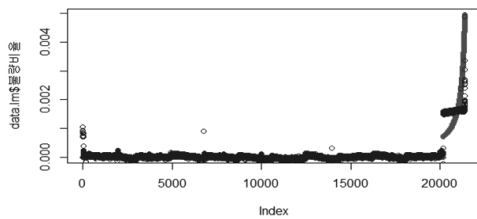
제약성의 문제로 결과를 얻지 못했다. 다만, A업체가 보유한 성형조건표 기준과 생산이력파일의 작업 설정치는 다소 차이를 보였다.

(A사 Injection Molding Chart Alternative code)

Experimental data	Actual data
C1	Cycle time
C2	Weighing time
C3	Injection time
C4	V-P position

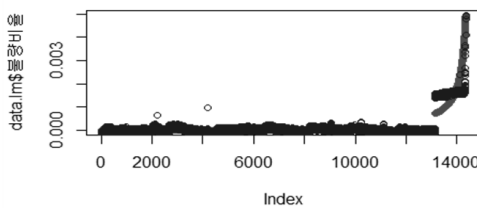
• 초기 단계

- 제조 공정상에서 수집한 2만 1천 2백 개의 데이터
- 불량률이 없는 제품의 제조 공정상에서 수집한 데이터 : 20,000개
- 불량률이 있는 제품의 제조 공정상에서 수집한 데이터 : 1,200개



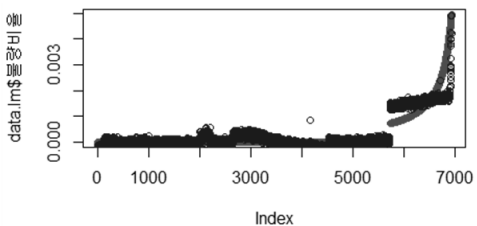
• 중간 과정(1) 단계

- 불량률이 없는 제품의 제조 공정상에서 수집한 데이터 : 15,000개
- 불량률이 있는 제품의 제조 공정상에서 수집한 데이터 : 1,200개



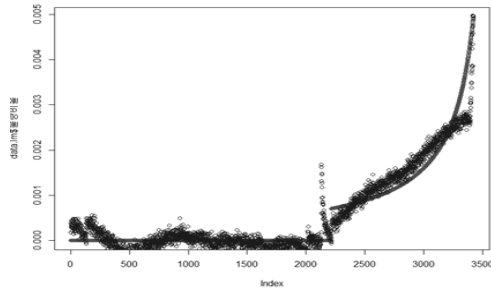
• 중간 과정(2) 단계

- 불량률이 없는 제품의 제조 공정상에서 수집한 데이터 : 8,000개
- 불량률이 있는 제품의 제조 공정상에서 수집한 데이터 : 1,200개



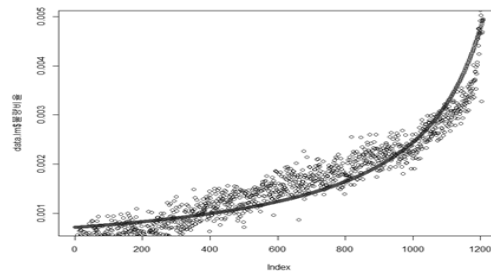
• 중간 과정(3) 단계

- 불량률이 없는 제품의 제조 공정상에서 수집한 데이터 : 2,000개
- 불량률이 있는 제품의 제조 공정상에서 수집한 데이터 : 1,200개



• 최종 단계

- 불량률이 없는 제품의 제조 공정상에서 수집한 데이터 : 0개
- 불량률이 있는 제품의 제조 공정상에서 수집한 데이터 : 1,200개



3.5 표준 TA를 활용한 불량 예측 사례

Data Set 신뢰성을 위하여 자동차 커넥트 사출 장비와 연계하여 표준 데이터 10 세트 추출 및 분석 API를 통하여 불량률을 분석하였으며, 업체의 실무 적용 결과를 통해 신뢰성을 검증하였다. 예측사례를 통하여 얻은 결과로 성형조건 설정치 중 설정항목은 15개 이상, 생산 부품에 영향을 줄 수 있는 설정 항목은 4개(사이클 타임, 사출시간, 금형온도, 원재료 수분 함량 등) 정도로 축소되었으며, 실제 불량률과 예측된 불량률이 유사한 단계, 예측하지 못한 측정 변수로 인한 불량 오차율, 설정 값과 이력파일이 관계성이 없는 실험 데이터도 측정이 되었다.

사출성형의 부품 성형 시간은 평균 사이클 타임 5~10초 정도이며, 금형 온도조절 모의실험 결과 제품생산 시간의 10%정도 단축 되었으나, A사 샘플 제품과 사출성형기의 냉각기 위치 조건에 따라 많은 차이를 보였다. 또한, 제조 업체의 정보 규정과 특성을 고려하여 한정된 자료를 제공받아서 실험 분석된 결과로 다양한 예측과 패턴모형은 다소 미흡한 점으로 나타났다.

산업유형별 데이터를 분석과 검증을 통해 산업유형별에 맞는 다양한 형태의 패턴을 분석할 수 있었다. Fig. 16(A~E)에 예측사례 분석 실험 데이터 및 그래프를 나타내었다.

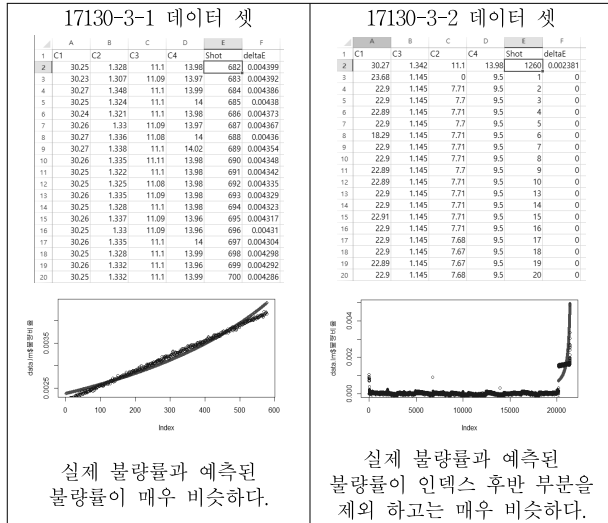


Fig. 16A. Case Studies Predict Experimental Data and Graphs

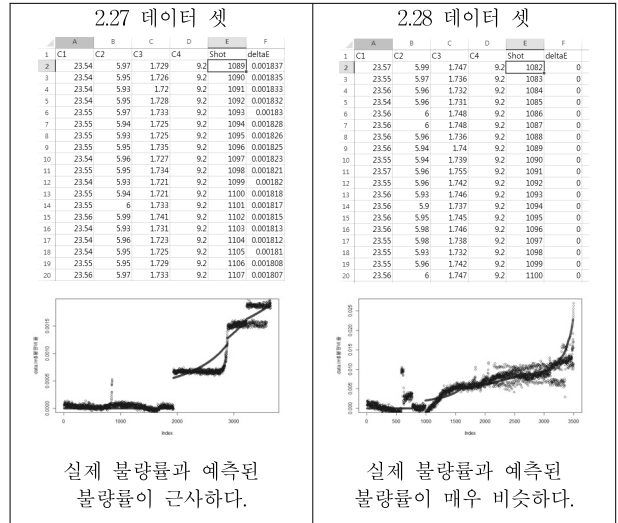


Fig. 16D. Case Studies Predict Experimental Data and Graphs

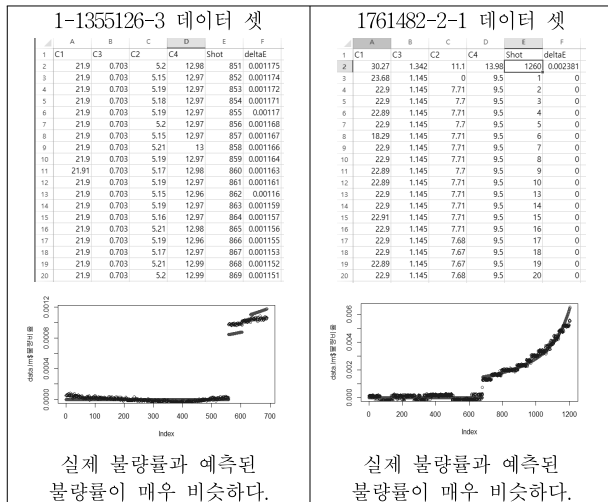


Fig. 16B. Case Studies Predict Experimental Data and Graphs

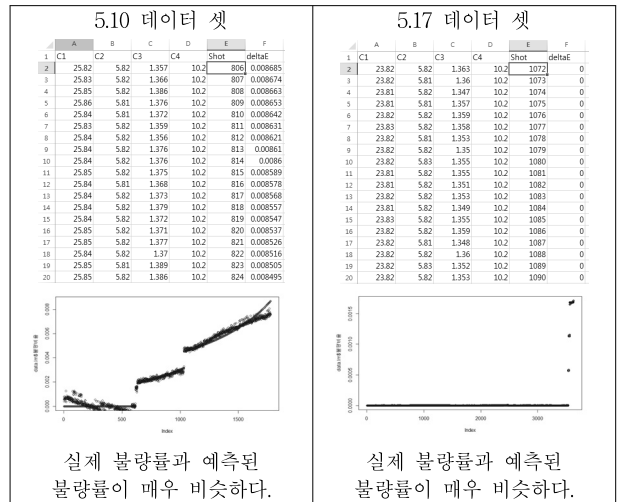
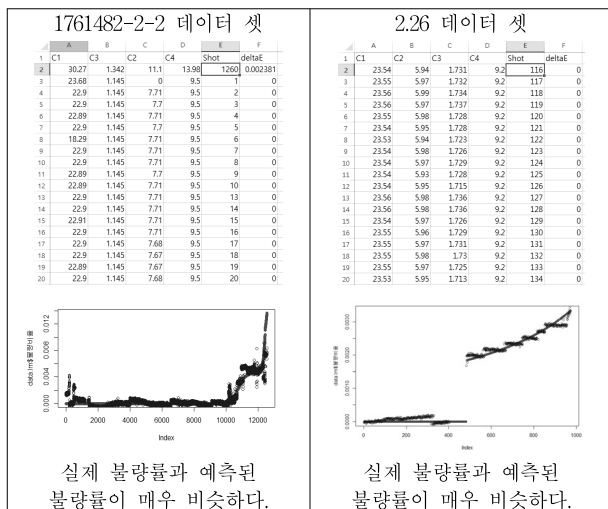


Fig. 16E. Case Studies Predict Experimental Data and Graphs



리즘, 시각화를 지원할 수 있는 저작도구(모니터링) 등에 대한 설계 및 개발 예정이다.

References

- [1] Jae Chun Kim and Seon-A Jin, "Information Visualization for the Manufacturing Process Optimization Based on Design of Experiment and Data Analysis," *KIPS Transactions on Software and Data Engineering*, Vol.4, No.9, pp.393-402, 2015.
- [2] Jae Chun Kim, Seon-A Jin, Young Hee Park, Seong Yeo Noh, and Hyun Dong Lee, "A Design for Realtime Monitoring System and Data Analysis Verification TA to Improve the Manufacturing Process using HW-SW Integrated Framework," *KIPS Transaction on Software and Data Engineering*, Vol.4, No.9, pp.357-370, 2015.
- [3] Hyun Sik Sim and Chang Ouk Kim, "Fault-Causing Process and Equipment Analysis of PCB Manufacturing Lines Using Data Mining Technique," *KIPS Transactions on Software and Data Engineering*, Vol.4, No.2, pp.65-70, 2015.
- [4] ARTIK 10 [Internet], <https://www.artik.io/modules/overview/artik-10>.
- [5] oneM2M [Internet], <http://www.onem2m.org/>.
- [6] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining: Concepts and Techniques," 3th ed., Acorn Publishingm, 2015.
- [7] Kyeongsoo Jeong, Byeonggon Kim, and Sangdo Jang, "A Study on the Development of Framework for Enhancing Data Quality in a Data Warehouse Environment," *Journal of Business Education*, Vol.19, pp.27-41, 1999.
- [8] Ian H. Witten, Eibe Frank, and Mark A. Hall, "Data Mining : Practical Machine Learning Tools and Techniques," 3th ed., Acorn Publishingm, Inc., 2013.
- [9] Myung-han Yu and Sangkyung Kim, "Improvement of SWoT-Based Real Time Monitoring System," *KIPS Transactions on Computer and Communication Systems*, Vol.4, No.7, pp.227-234, 2015.



정민승

e-mail : jms4112@naver.com

2016년 경남정보대학교 컴퓨터정보공학과 (학사)

2014년~현 재 ㈜나라시스템 연구원

2016년~현 재 부경대학교 컴퓨터공학과 석사과정

관심분야: 부실채권 관리, 데이터마이닝, 프레임워크, 포렌식



진선아

e-mail : mystarta@hanmail.net

2010년 부경대학교 컴퓨터멀티미디어공학 (학사)

2015년~현 재 ㈜나라시스템 연구원

2016년~현 재 한양사이버대학교

ITMBA 석사과정

관심분야: 부실채권 관리, 데이터마이닝, 프레임워크, 포렌식



조우현

e-mail : whcho@pknu.ac.kr

1985년 경북대학교 전자공학과

전산공학전공(학사)

1988년 경북대학교 전자공학과(석사)

1998년 경북대학교 전자공학과

전산공학전공(박사)

1989년~현 재 부경대학교 컴퓨터공학과 교수

관심분야: 지능형 데이터베이스, 멀티미디어 인덱싱,

객체 데이터베이스 관리 기술