

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2020-0010671
(43) 공개일자 2020년01월31일

(51) 국제특허분류(Int. Cl.)
G05B 23/02 (2006.01) G06N 3/08 (2006.01)
(52) CPC특허분류
G05B 23/0243 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2018-0078640
(22) 출원일자 2018년07월06일
심사청구일자 없음

(71) 출원인
에임시스템 주식회사
경기도 군포시 엘에스로 175, 6층 (산본동, 에스
에이타워)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암
동5가)
(72) 발명자
안민정
서울특별시 관악구 남현8길 8-6
이홍철
서울특별시 서초구 동광로19길 51, 701호
(뒷면에 계속)
(74) 대리인
윤재승

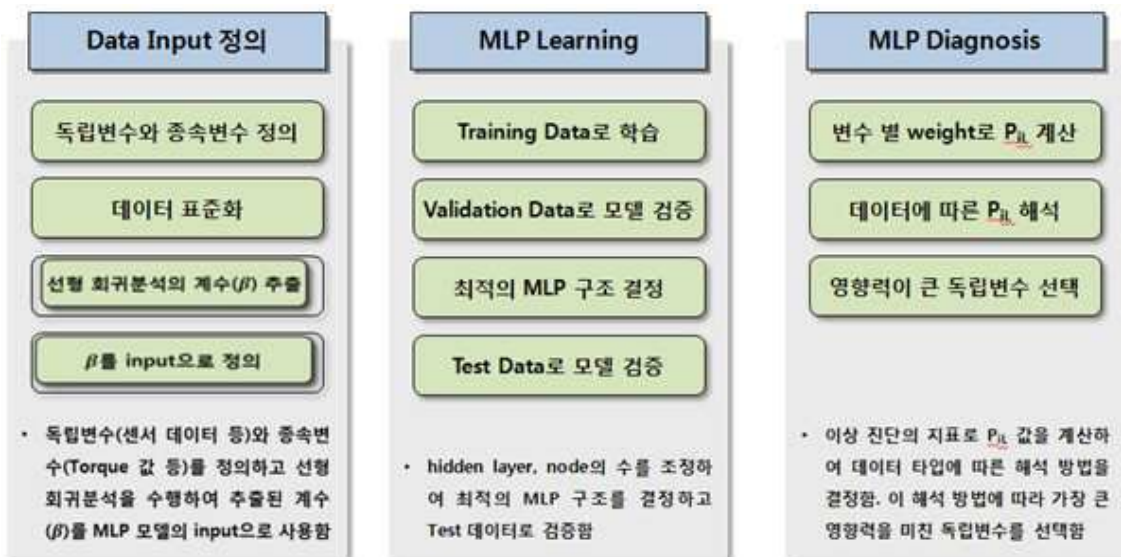
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 기계 학습 기반의 설비 이상 진단 시스템 및 방법

(57) 요약

본 발명에 따른 기계 학습 기반의 설비 이상 진단 방법은, Regression weight를 입력으로 하는 다층신경망(Multilayer Perceptron) 모델을 학습시켜 추출된 weight로 이상 진단 지표 P_{IL} 을 계산하여 이상의 원인이 되는 변수의 우선순위를 산출하는, MLP 모델을 이용한 상기 이상 진단 과정을 포함한다.

대표도 - 도33



(72) 발명자

최영진

경상북도 구미시 백산로 153 동양한신아파트 102동
601호

홍진석

서울특별시 동대문구 왕산로 86 A동 1702호

권상현

서울특별시 서초구 방배로3길 10-3 101동 408호

김명소

서울특별시 동대문구 무학로49길 22 404호

곽유영

경기도 고양시 덕양구 충장로152번길 39
햇빛마을20단지아파트 2003동 806호

변성준

경기도 성남시 분당구 정자일로 121, 101동 2601호

명세서

청구범위

청구항 1

기계 학습 기반의 설비 이상 진단 방법으로서,

Regression weight를 입력으로 하는 다층신경망(Multilayer Perceptron) 모델을 학습시켜 추출된 weight로 이상 진단 지표 P_{il} 을 계산하여 이상의 원인이 되는 변수의 우선순위를 산출하는, MLP 모델을 이용한 상기 이상 진단 과정을 포함하는, 기계 학습 기반의 설비 이상 진단 방법.

청구항 2

제1항에 있어서,

MLP 모델을 이용한 상기 이상 진단 과정은,

독립변수와 종속변수를 정의하고, 데이터 표준화를 수행하고, 선형 회귀분석의 계수 β 를 추출하고, β 를 input으로 정의하는 데이터 입력 정의 과정;

Training Data로 학습하고, Validation data로 모델을 검증하고, 최적의 MLP 구조를 결정하고, Test Data로 모델을 검증하는 MLP 학습 과정; 및

변수 별 weight로 P_{il} 을 계산하고, 데이터에 따라 P_{il} 을 해석하고, 영향력이 큰 독립변수를 선택하는 MLP 진단 과정을 포함하는, 기계 학습 기반의 설비 이상 진단 방법.

청구항 3

제2항에 있어서,

상기 데이터 입력 정의 과정은, 독립변수와 종속변수를 정의하고 선형 회귀분석을 수행하여 추출된 계수 β 를 MLP 모델의 input으로 사용하는, 기계 학습 기반의 설비 이상 진단 방법.

청구항 4

제2항에 있어서,

상기 MLP 학습 과정은, hidden layer, node의 수를 조정하여 최적의 MLP 구조를 결정하고 Test data로 검증하는, 기계 학습 기반의 설비 이상 진단 방법.

청구항 5

제2항에 있어서,

상기 MLP 진단 과정은, 이상 진단의 지표로 P_{il} 값을 계산하여 데이터 타입에 따른 해석 방법을 결정하는, 기계 학습 기반의 설비 이상 진단 방법.

발명의 설명

기술 분야

[0001] 본 발명은 기계 학습 기반의 설비 이상 진단 시스템 및 방법에 관한 것으로, 보다 상세하게는 데이터 독립변수 간의 Regression Weight 계수값을 입력으로 하여 MLP 모델을 학습시켜 추출된 지표를 통해 이상의 원인이 되는 변수의 우선순위를 산출하는 시스템 및 방법에 관한 것이다.

배경 기술

[0002] 설비 시스템에 결함이나 손상이 발생할 경우, 시스템의 가동 중단을 야기할 수 있으며 이로 인한 사회적, 경제

적 피해의 규모는 막대하다. 따라서 고장이 발생하기 전에 예방정비를 실행하는 것이 중요하며 현재 대부분의 시스템은 일정한 시간 간격을 두고 실시하는 정기적 예방정비에 의존하고 있다. 그러나 정기적 예방정비는 부품의 실제 결함 유무와 관계없이 일정한 주기에 의해 실시되므로 정상 부품의 불필요한 교체로 인한 비용 손실을 발생시키고 갑작스런 시스템의 고장을 예방하는 데에 한계가 있다. 따라서 이러한 문제들을 해결하기 위해 최근 고장예지 및 건전성관리(PHM: prognostics and health management)기술이 활발히 연구되고 있다. 고장예지 및 건전성관리 기술은 설비 시스템의 상태를 실시간으로 감시하여 시스템의 이상을 조기에 감지하고 미래에 발생할 고장을 미리 예측한다. 따라서 적절한 조치를 미리 취함으로써 불필요한 유지보수 비용을 줄일 수 있고, 시스템의 안전성과 신뢰성을 높여 재앙적인 사고를 예방할 수 있다.

[0003] 고장예지 및 건전성관리 기술에는 크게 두 가지 방법이 있다. 첫째로, 데이터주도방법(datadriven approach)은 데이터를 이용하여 시스템의 신뢰성, 건전성 정보를 통계적 방법으로 추론하는 것이다. 기계학습(machine learning)기법은 가장 많이 사용되는 기법으로 건전성인자와 고장 여부의 관계를 훈련하여 미래고장을 예측할 수 있다. 이 방법은 물리적 손상 모델을 구현하기 어려운 다변량 시스템에 적용이 가능하다는 장점이 있지만, 훈련을 위해 많은 양의 데이터가 필요하다. 다음으로 모델기반방법(model based approach)은 물리적 고장 모델을 기반으로 고장을 진단 및 예지한다. 이 방법은 정확도가 높으며 적은 양의 고장 데이터로 고장 진단이 가능하다. 모델의 변수를 바꿔줌으로써 다양한 운행 환경에서도 적용 가능하다. 그러나, 고장 메커니즘 파악이 어렵거나 모델 변수의 수가 매우 많은 경우 모델이 실제 고장 메커니즘을 온전히 구현하지 못하므로 적용 분야가 한정적이다.

[0004] 공정 이상치 탐지 및 분류 그리고 진단에 대한 기술은 SPC 차트, k-NN, 데이터마이닝 기법, 인공신경망 등과 같은 다양한 방법론의 기술들이 연구되었고 진행되고 있다. 현 공정에서는 대표적인 공정관리도 방법으로써 SPC차트를 많이 사용한다. 하지만 SPC 차트는 관리한계도가 높아지는 현상 때문에 이상치에 대해 효과적으로 탐지하기 어려웠고 판정의 어려움으로 실 공정에서 나오는 데이터의 정상/비정상에 대한 판단여부도 불확실한 점이 있었다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적 과제는 데이터 독립변수간의 Regression Weight 계수값을 입력으로 하여 MLP 모델을 학습시켜 추출된 지표를 통해 이상의 원인이 되는 변수의 우선순위를 산출할 수 있는 기계학습 기반의 설비 이상 진단 시스템 및 방법을 제공하는데 있다.

과제의 해결 수단

[0006] 상기 기술적 과제를 해결하기 위한 본 발명에 따른 기계 학습 기반의 설비 이상 진단 방법은, Regression weight를 입력으로 하는 다층신경망(Multilayer Perceptron) 모델을 학습시켜 추출된 weight로 이상 진단 지표 P_{il} 을 계산하여 이상의 원인이 되는 변수의 우선순위를 산출하는, MLP 모델을 이용한 상기 이상 진단 과정을 포함한다.

[0007] MLP 모델을 이용한 상기 이상 진단 과정은, 독립변수와 종속변수를 정의하고, 데이터 표준화를 수행하고, 선형 회귀분석의 계수 β 를 추출하고, β 를 input으로 정의하는 데이터 입력 정의 과정; Training Data로 학습하고, Validation data로 모델을 검증하고, 최적의 MLP 구조를 결정하고, Test Data로 모델을 검증하는 MLP 학습 과정; 및 변수 별 weight로 P_{il} 을 계산하고, 데이터에 따라 P_{il} 을 해석하고, 영향력이 큰 독립변수를 선택하는 MLP 진단 과정을 포함할 수 있다.

[0008] 상기 데이터 입력 정의 과정은, 독립변수와 종속변수를 정의하고 선형 회귀분석을 수행하여 추출된 계수 β 를 MLP 모델의 input으로 사용할 수 있다.

[0009] 상기 MLP 학습 과정은, hidden layer, node의 수를 조정하여 최적의 MLP 구조를 결정하고 Test data로 검증할 수 있다.

[0010] 상기 MLP 진단 과정은, 이상 진단의 지표로 P_{il} 값을 계산하여 데이터 타입에 따른 해석 방법을 결정할 수 있다.

발명의 효과

[0011] 상기된 본 발명에 의하면, 데이터 독립변수간의 Regression Weight 계수값을 입력으로 하여 MLP 모델을 학습시켜 추출된 지표를 통해 이상의 원인이 되는 변수의 우선순위를 산출할 수 있는 기계학습 기반의 설비 이상 진단 시스템 및 방법을 제공할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 기계 학습 기반의 설비 이상 탐지/분류/진단 시스템의 전체적인 구성을 나타낸다.

도 2는 본 발명의 일 실시예에 따른 기계 학습 기반의 설비 이상 탐지/분류/진단 시스템의 전체 공정을 나타낸다.

도 3은 데이터 전처리 과정의 일 예를 보여준다.

도 4는 CVD(Chemical vapor Deposition) 설비와 스톡커(Stocker) 설비로부터 수집되는 데이터의 일 예를 보여준다.

도 5는 차원 축소 과정을 설명하는 도면이다.

도 6은 차원 축소 결과로서, 1차 차원 축소의 예를 보여준다.

도 7은 차원 축소 결과로서, 2차 차원 축소의 예를 보여준다.

도 8은 안정화 구간을 설정하는 과정의 일 예를 보여준다.

도 9는 특질 추출의 일 예를 보여준다.

도 10은 이상 탐지 과정의 일 예를 보여준다.

도 11은 SVR 모델을 적용하여 이상을 탐지하는 과정을 나타낸다.

도 12는 SVR 학습 과정에서 선택할 수 있는 SVR 커널들의 예를 보여준다.

도 13은 Golden Line과 Margin의 예를 보여준다.

도 14는 SVR 예측 결과 이상으로 탐지된 데이터의 예를 보여준다.

도 15는 k-NNDD 모델을 적용하여 이상을 탐지하는 과정을 나타낸다.

도 16은 DTW 알고리즘의 개념을 나타내는 도면이다.

도 17은 계층적 군집화를 통해 새로운 관측치로부터 정상 관측치가 얻어지는 것을 나타낸 도면이다.

도 18은 비모수 추정 방법을 활용한 k-NNDD의 이상치 탐지 기법의 개념을 나타낸다.

도 19는 SSE로 Elbow point를 선택하여 군집수를 결정하는 것과 DTW 거리를 기반으로 한 계층적 군집화 결과를 나타낸다.

도 20은 SSE와 실루엣 통계량의 개념을 나타낸다.

도 21은 DTW distance를 기반으로 한 계층적 군집화 결과를 나타낸다.

도 22는 기울기를 대표 값으로 k-means를 수행한 분리형 군집화 결과를 나타낸다.

도 23은 이상 분류 라벨링 결과(3 class)를 나타낸다.

도 24는 k-NNDD 모델을 통해 novelty score를 계산하는 것을 나타낸다.

도 25는 k-NNDD 모델을 이용한 이상 분류 과정을 나타낸다.

도 26은 k-NNDD 모델 분류 결과의 예를 나타낸다.

도 27은 RNN-LSTM 모델의 개념을 나타낸다.

도 28은 RNN-LSTM 모델을 통해 이상을 분류하는 과정을 나타낸다.

도 29는 RNN-LSTM 모델의 Accuracy와 Cost의 예를 보여준다.

도 30은 MLP 모델의 다층신경망 모델을 나타낸다.

도 31은 데이터셋에 따라 가중치를 반영하여 k-NNDD, RNN, MLP 모델을 선정하는 것의 예를 나타낸다.

도 32는 MLP 모델을 통해 이상 진단 지표 P_{il} 을 계산하는 과정을 나타낸다.

도 33은 MLP 모델을 이용한 이상 진단 과정을 나타낸다.

도 34는 MLP 모델 적용 결과의 예를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0013] 이하에서는 도면을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다. 이하 설명 및 첨부된 도면들에서 실질적으로 동일한 구성요소들은 각각 동일한 부호들로 나타냄으로써 중복 설명을 생략하기로 한다. 또한 본 발명을 설명함에 있어 관련된 공지기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그에 대한 상세한 설명은 생략하기로 한다.
- [0014] 도 1은 본 발명의 일 실시예에 따른 기계 학습 기반의 설비 이상 탐지/분류/진단 시스템의 전체적인 구성을 나타낸다.
- [0015] 본 발명의 일 실시예에 따른 설비 이상 탐지/분류/진단 시스템은 설비의 이상을 실시간으로 탐지하고 분류한 후 이상이 발생한 원인을 진단하는 시스템으로서, 데이터 전처리, 특징 추출, 이상 탐지, 이상 분류 라벨링, 이상 분류, 이상 진단의 과정을 포함한다.
- [0016] 도 2는 본 발명의 일 실시예에 따른 기계 학습 기반의 설비 이상 탐지/분류/진단 시스템의 전체 공정을 나타낸다.
- [0017] 본 발명의 일 실시예에 따른 설비 이상 탐지/분류/진단 시스템은 설비 타입 별로 분석 모델을 선택하여 적용할 수 있는 유연성 있고 확장 가능한 프로세스로 구성된다. 또한 이상 탐지, 분류, 진단 과정은 각 모델 별로 학습(learning)과 예측(prediction)의 두 페이스(phase)로 구성된다.
- [0018] 데이터 전처리 과정은 데이터 수집, 차원 축소, 안정화구간 설정 과정을 포함한다.
- [0019] 특징 추출 과정은 데이터 영역 별 지표 추출 과정을 포함한다.
- [0020] 이상 탐지 과정은 SVR 모델 적용, k-NNDD 모델 적용 과정을 포함한다.
- [0021] 이상 분류 라벨링 과정은 군집 수 결정 및 군집화 과정을 포함한다.
- [0022] 이상 분류 과정은 k-NNDD 모델 적용, RNN-LSTM 모델 적용, MLP 모델 적용 과정을 포함한다.
- [0023] 이상 진단 과정은 MLP 모델 적용 과정을 포함한다.
- [0024] 도 3은 데이터 전처리 과정의 일 예를 보여준다.
- [0025] 데이터 전처리 과정에서는, 설비에서 수집되는 대량의 계측 데이터를 수집하여 유의미한 변수를 추출하고, 해당 변수의 데이터 내에서 안정화 구간을 사전 정의된 규칙에 의해 자동으로 설정한다.
- [0026] 도 4는 CVD(Chemical vapor Deposition) 설비와 스톡어(Stocker) 설비로부터 수집되는 데이터의 일 예를 보여준다. 설비 타입에 따라 저장되는 데이터의 형태와 안정화 구간에 대한 정의가 다르기 때문에, 데이터 수집 과정에서 설비 별 초기 데이터를 로딩하기 위한 프로그래밍이 필요하다.
- [0027] 도 5는 차원 축소 과정을 설명하는 도면이다.
- [0028] 변수가 많을 경우 모델을 학습하고 예측하는 연산의 계산 복잡도가 증가하고, 노이즈로 인한 성능 손실 발생할 수 있으므로 비슷한 결과를 산출할 수 있는 최소의 변수를 선택한다. 변수 선택은 도메인 지식이 풍부한 엔지니어의 정성적 판단이 가장 중요하다. 정성적 판단이 어려울 경우 표준화 프로세스(중속변수 제거 및 특이값 분해)를 적용하여 정량적 기준으로 변수를 선택할 수 있다.
- [0029] 도 5를 참조하면, 1차로, 전체 데이터에서 중속 변수를 제거하여 차원을 축소한다.
- [0030] VIF 값은 다음 수학적식으로 표현된다.

수학식 1

$$VIF = \frac{1}{1 - R_k^2}$$

[0031]

[0032]

VIF(Variance Inflation Factor)는 회귀분석에서 변수들 간 독립성을 확인하기 위해 쓰이는 지표로 해당 변수의 VIF 값이 높을수록 다른 독립변수들과 상관관계가 높음을 의미한다.

[0033]

여기서, 단계적 VIF 방법을 사용할 수 있다. 즉, 1. 각 변수들의 VIF 값을 모두 측정하고; 2. 가장 큰 VIF 값을 가진 변수를 제거하고; 3. 남은 변수들의 VIF 값을 측정하고; 4. 2,3 과정을 모든 변수가 임계값(예컨대 10) 이하가 될 때까지 반복하여 최종적으로 남은 변수를 선택한다.

[0034]

2차로, 특이값 분해 과정을 통해, 종속변수가 제거된 데이터에서 차원을 축소한다. SVD(Singular Value Decomposition)에서, 특이값 분해는 행렬 A를 회전부분 V, U와 증폭부분 Σ (특이값)로 분해하여 Σ 의 원소들 중 큰 값들을 추출하여 차원을 축소한다. 이때 Truncated SVD 방법을 사용할 수 있다. A'와 코사인 유사도가 가장 높은 X들로 최종 변수를 선택한다.

[0035]

도 6은 차원 축소 결과로서, 1차 차원 축소의 예를 보여준다. 도 6을 참조하면, VIF 테스트로 171개 변수 중 49개 변수를 추출하였다.

[0036]

도 7은 차원 축소 결과로서, 2차 차원 축소의 예를 보여준다. 도 7을 참조하면, SVD 테스트로 49개 변수 중 40개 변수를 추출하였다. 2차 차원 축소를 통해 유의미한 변수가 추출되었다.

[0037]

도 8은 안정화 구간을 설정하는 과정의 일 예를 보여준다.

[0038]

안정화 구간 설정 과정에서, 계측 데이터의 유형에 따라 주요 관리 대상이 되는 안정화 구간의 정의는 다르나 일반화시킬 수 있는 기준을 적용하여 안정화 구간을 설정할 수 있다.

[0039]

구체적으로, 전체 데이터를 X-bar Chart에 입력하여 UCL, LCL값을 추출하고, 관리도 내에 처음(다만 5회 연속 등으로 수정 가능)으로 데이터가 들어오는 지점을 안정화 구간의 시작으로 설정하고, 한 공정신호의 영역을 시작, 안정화구간, 끝의 세 스텝으로 구분한다.

[0040]

도 9는 특질 추출의 일 예를 보여준다.

[0041]

데이터 영역 별 지표 추출로서, 선택한 변수의 계측 데이터는 시계열 데이터이므로 다양한 분석 모델을 적용하기 위해 한 사이클의 계측 데이터를 대표할 수 있는 특질(설비 타입에 따라 중요도가 다를 수 있음)을 추출한다.

[0042]

이상 탐지 과정은 이상 분류와 진단을 위해 대량의 데이터로부터 이상 탐지 모델을 생성하여 이상 데이터를 추출하는 과정이다.

[0043]

구체적으로, One class classification 중 경계선을 통한 분류 방법(boundary method)인 SVR과 거리기반의 Novelty detection 방법인 k-NNDD 모델을 적용하여 공통적으로 탐지된 이상데이터를 수집하여 라벨링한다. 이때 0은 정상율, 1은 이상을 나타낸다.

[0044]

도 10은 이상 탐지 과정의 일 예를 보여준다.

[0045]

이상 탐지 과정에서는 SVR 모델을 적용한다. 정상 상태의 주기신호에 대한 평균을 나타는 함수(Golden Line)를 추정하기 위하여 비선형 regression 기법인 SVR 모델을 적용한다. 추정된 함수에 관리한계선(Margin)을 설정하고 이 경계를 벗어난 데이터(Outlier)의 비율(예컨대 Threshold, 10% 등)로 이상 여부를 판단할 수 있다.

[0046]

도 11은 SVR 모델을 적용하여 이상을 탐지하는 과정을 나타낸다.

[0047]

데이터 전처리 과정에서 데이터 표준화를 수행한다. SVR 모델이 다양한 타입의 계측 데이터에 범용적으로 적용될 수 있도록 데이터를 표준화한다.

[0048]

SVR 학습 과정에서, SVR 커널을 선택하고, 커널 별 파라미터를 설정하고, Golden Line을 설정하고, Margin을 설정한다. 이때 데이터에 따라 적합한 SVR 커널을 선택하고 데이터를 학습하여 정상 데이터의 대표 함수와 관리한

계선(Margin)을 설정한다.

- [0049] SVR 예측 과정에서, 이상 데이터 수를 추출하고 Threshold 기준 이상을 탐지한다. 학습 단계에서 설정된 관리한 계선(Margin)에서 벗어나는 데이터의 개수를 수집하여 사용자 정의된 Threshold를 기준으로 이상 여부를 판단하여 라벨링한다.
- [0050] 도 12는 SVR 학습 과정에서 선택할 수 있는 SVR 커널들의 예를 보여준다. Laplacian Kernel은 Parameter가 분산이므로 비교적 처리 속도가 빠르고 시계열 데이터 특성에 적합하다.
- [0051] SVR 학습 과정에서, Golden Line 설정 시, N개 cycle의 SVR-line (Laplacian Kernel)의 Median 값을 Golden Line으로 정의할 수 있다.
- [0052] Margin 설정 시, N초 별 SVR 값에 Whiskers (IQR, Interquartile Range에 1.5를 곱한 값) 방법을 적용하여 Margin으로 설정하고 이 경계를 벗어나는 데이터를 이상(비율로 설정)으로 탐지할 수 있다.
- [0053] 도 13은 Golden Line과 Margin의 예를 보여준다.
- [0054] SVR 예측 과정에서, 학습 단계에서 설정된 Margin을 벗어나는 데이터를 이상(비율로 설정, default : 10%)으로 탐지할 수 있다.
- [0055] 도 14는 SVR 예측 결과 이상으로 탐지된 데이터의 예를 보여준다.
- [0056] 또한 이상 탐지 과정에서는 k-NNDD (k-Nearest Neighbor Data Description) 모델을 적용한다.
- [0057] k-NNDD 모델은 거리 기반의 이상 탐지 방법으로 novelty score를 계산하여 정상 집단에 속하는지 여부를 판단한다. Novelty score란 새로운 관측치와 이웃패턴 간의 거리를 비율적으로 계산한 분류 척도이며, 새로운 관측치와 이웃 패턴간의 거리가 멀수록 이상치로 분류할 확률을 높게 추정한다.
- [0058] 도 15는 k-NNDD 모델을 적용하여 이상을 탐지하는 과정을 나타낸다.
- [0059] 데이터 전처리 과정에서 DTW distance를 계산한다. DTW distance는 k-NNDD 모델에 사용할 거리의 종류 중 하나이다. 설비의 타입에 따라 적합한 거리의 종류를 결정할 수 있다.
- [0060] k-NNDD 학습 과정에서, 계층적 군집화를 수행하고, 정상 모델을 설정하고, k와 Threshold를 결정한다. 계층적 군집화를 통해 주어진 학습 데이터에서 정상 모델을 분류할 수 있고, 데이터 유형 별로 시행착오 테스트를 통해 k(비교할 이웃 데이터 수)와 Threshold(이상 탐지 임계값)를 결정할 수 있다.
- [0061] k-NNDD 예측 과정에서, Novelty score를 계산하고, 이상을 탐지한다. 학습 단계에서 설정된 k개의 이웃 데이터로 Novelty score를 계산하여 Threshold를 초과할 경우 이상으로 탐지한다.
- [0062] DTW(Dynamic Time Warping) 알고리즘은 두 개의 시계열 데이터 간에 유사도를 알아내기 위한 알고리즘 중 하나로 동적 패턴의 시변성을 반영한다. DTW 알고리즘의 일반적인 계산 복잡도는 $O(N^2)$ 으로 대용량 트랜잭션에 효율적이다.
- [0063] 도 16은 DTW 알고리즘의 개념을 나타내는 도면이다.
- [0064] DTW distance로 계층적 군집화(Hierarchical clustering)를 통해 주어진 학습 데이터에서 정상 모델을 분류할 수 있다.
- [0065] 도 17은 계층적 군집화를 통해 새로운 관측치로부터 정상 관측치가 얻어지는 것을 나타낸 도면이다.
- [0066] k-NNDD(k-Nearest Neighbor Data Description) 알고리즘은 One class classification 중 경계선을 통한 분류 방법(boundary method)으로 인접한 관측치 사이의 거리를 기반으로 이상 여부를 판단한다. K-FDC 방법론에서는 DTW distance를 기준 거리로 사용한다.
- [0067] 도 18은 비모수 추정 방법을 활용한 k-NNDD의 이상치 탐지 기법의 개념을 나타낸다.
- [0068] 다음으로, 이상 분류 라벨링 과정을 설명한다.
- [0069] 이상 분류에 사용할 k-NNDD 모델과 RNN-LSTM 모델은 사전에 정의된 분류 class가 있어야 하는(Labeling) 지도 학습 (Supervised Learning) 방법이므로, 비지도 학습(Unsupervised Learning) 방법인 군집화(Clustering)를 통해 앞 단계(SVR, k-NNDD)에서 탐지한 이상 데이터의 분류 class를 생성한다. 최적의 군집 수는 군집화 평가지표(SSE, 실루엣 통계량 등)를 도식화하여 지표에 따른 Elbow point로 결정한다. 결정된 군집 수(K)로 군집화

(Clustering)하여 1-K로 한 공정 사이클 데이터를 Labeling 한다. 군집화 방법은 한 공정 사이클 데이터의 DTW distance, 평균값, 최대값, 중앙값 등 사용하는 대표 값에 적합한 방법(계층적 군집화, 분리형 군집화 등)을 테스트를 통해 선정할 필요가 있다.

- [0070] 도 19는 SSE로 Elbow point를 선택하여 군집수를 결정하는 것과 DTW 거리를 기반으로 한 계층적 군집화 결과를 나타낸다.
- [0071] 군집 수 결정에서, 군집 성능 평가 지표(SSE, 실루엣 통계량 등)를 도식화하여 최적의 군집 수를 선택한다.
- [0072] 도 20은 SSE와 실루엣 통계량의 개념을 나타낸다.
- [0073] 도 21은 DTW distance를 기반으로 한 계층적 군집화 결과를 나타낸다.
- [0074] 도 22는 기울기를 대표 값으로 k-means를 수행한 분리형 군집화 결과를 나타낸다.
- [0075] 도 23은 이상 분류 라벨링 결과(3 class)를 나타낸다.
- [0076] 다음으로, 이상 분류 과정을 설명한다.
- [0077] 이상 분류 과정에서, k-NNDD(k-Nearest Neighbor Data Description) 모델을 적용한다. k-NNDD 모델은, 새로운 관측치가 있을 때, 라벨링된 이상 분류 모델에 대해 class 별로 k-NNDD를 적용하여 novelty score($d1/d2$)값이 가장 작은 순서로 해당 class에 속할 우선순위를 부여한다.
- [0078] 도 24는 k-NNDD 모델을 통해 novelty score를 계산하는 것을 나타낸다.
- [0079] 도 25는 k-NNDD 모델을 이용한 이상 분류 과정을 나타낸다.
- [0080] 데이터 전처리 과정에서, DTW distance matrix를 계산한다. 즉, 이상 종류 별로 Labeling 된 모델 내 DTW distance matrix를 계산한다.
- [0081] k-NNDD 학습 과정에서 k 값을 결정하는데, k 값(인접한 관측치의 개수)은 크기가 최소인 모델 크기의 예컨대 20%(최소 5 이상)를 기준으로 할 수 있다.
- [0082] k-NNDD 예측 과정에서, 학습 단계에서 설정된 k 개의 이웃 데이터로 Novelty score를 계산하여 가장 작은 값을 가지는 순서로 class의 우선순위를 부여한다.
- [0083] k-NNDD 모델 분류 결과는, 새로운 관측치가 있을 때, 학습된 이상 분류 모델에 대해 class 별로 k-NNDD를 적용하여 novelty score($d1/d2$)값이 가장 작은 순서로 해당 class에 속할 우선순위를 부여한다.
- [0084] 도 26은 k-NNDD 모델 분류 결과의 예를 나타낸다.
- [0085] 또한 이상 분류 과정에서, RNN-LSTM 모델을 적용한다. 즉, Deep Learning 모델 중 순차적인 정보를 처리하는데 적합한 RNN-LSTM(Recurrent Neural Network-Long Short Term Memory) 모델에 Softmax 함수를 조합하여 이상을 분류한다.
- [0086] 도 27은 RNN-LSTM 모델의 개념을 나타낸다.
- [0087] RNN-LSTM 모델에서, 라벨링된 데이터를 훈련/검증/테스트 셋으로 나누어 RNN-LSTM 모델을 훈련시키고 성능을 검증한 후 실시간으로 데이터를 분류한다. 이때 반복 학습을 통해 Hyper parameter를 조정하는 작업이 중요하다.
- [0088] 도 28은 RNN-LSTM 모델을 통해 이상을 분류하는 과정을 나타낸다.
- [0089] Hyper parameter 설정 과정에서, RNN-LSTM 구조를 형성하고, Training 방안을 구성한다. 여기서, 최적의 RNN-LSTM 구조를 결정하기 위하여 Cell stack size, hidden layer, put size 등의 parameter를 설정한다. 학습을 위한 parameter로는 batch size, Hidden layer, learning rate, Max iteration, drop out 등이 있다.
- [0090] RNN-LSTM 학습 과정에서, Training data로 학습을 수행하고, Validation Data로 과적합을 체크하고, 최적의 RNN-LSTM 구조를 확정한다. 이때 초기 구성한 Hyper parameter를 학습과 검증을 통해 조정하여 최적의 RNN-LSTM 구조를 확정한다.
- [0091] RNN-LSTM 예측 과정에서, Test data로 모델을 검증하고, 실시간 이상 분류를 수행한다. 즉, Test 데이터를 바탕으로 본 모델의 성능을 검증한 후 실시간 데이터로 분류한다.
- [0092] RNN-LSTM 모델의 성능은, 학습 횟수가 증가할수록 Cost는 감소하고 정확도는 증가한다. Cost는 학습을 통해 줄

이고자 하는 오차의 전체 합 / Batch Size 이고, Accuracy는 예측 Class와 실제 Class와의 정확도이다.

[0093] 도 29는 RNN-LSTM 모델의 Accuracy와 Cost의 예를 보여준다.

[0094] 또한 이상 분류 과정에서, MLP(Multilayer Perceptron) 모델을 적용한다. MLP 모델은 Regression weight를 입력으로 하는 다층신경망(Multilayer Perceptron) 모델을 학습시켜 이상을 분류한다.

[0095] 도 30은 MLP 모델의 다층신경망 모델을 나타낸다.

[0096] 이상 분류 종합 과정에서, k-NNDD, RNN, MLP 모델로 분류한 결과에 사용자 정의 가중치(1:2:1 등)를 반영하여 최종적인 이상 분류 class를 확정한다. 이때 데이터 타입에 따라 적합한 분류 모델을 선정하는 테스트가 필요하다.

[0097] 도 31은 데이터셋에 따라 가중치를 반영하여 k-NNDD, RNN, MLP 모델을 선정하는 것의 예를 나타낸다.

[0098] 다음으로, 이상 진단 과정을 설명한다.

[0099] 이상 진단 과정에서는, MLP(Multilayer Perceptron) 모델을 적용한다. Regression weight를 입력으로 하는 다층신경망(Multilayer Perceptron) 모델을 학습시켜 추출된 weight로 이상 진단 지표 P_{il} 을 계산하여 이상의 원인이 되는 변수의 우선순위를 산출한다.

[0100] 도 32는 MLP 모델을 통해 이상 진단 지표 P_{il} 을 계산하는 과정을 나타낸다.

[0101] MLP 모델은 인공신경망을 이용한 이상 진단 기법으로 변수의 weight 값을 조합한 지표인 P_{il} 값을 계산하여 해당 cycle이 특정 군집으로 분류되는데 가장 크게 기여한 변수를 찾아낸다. P_{il} 값이 높을수록 해당 변수의 영향력이 크다고 볼 수 있으며 다른 군집 cycle의 패턴과 차이를 보인다.

[0102] 도 33은 MLP 모델을 이용한 이상 진단 과정을 나타낸다.

[0103] 데이터 입력 정의 과정에서, 독립변수와 종속변수를 정의하고, 데이터 표준화를 수행하고, 선형 회귀분석의 계수(β)를 추출하고, β 를 input으로 정의한다. 즉, 독립변수(센서 데이터 등)와 종속변수(Torque 값 등)를 정의하고 선형 회귀분석을 수행하여 추출된 계수(β)를 MLP 모델의 input으로 사용한다.

[0104] MLP 학습 과정에서, Training Data로 학습하고, Validation data로 모델을 검증하고, 최적의 MLP 구조를 결정하고, Test Data로 모델을 검증한다. 여기서, hidden layer, node의 수를 조정하여 최적의 MLP 구조를 결정하고 Test data로 검증한다.

[0105] MLP 진단 과정에서, 변수 별 weight로 P_{il} 을 계산하고, 데이터에 따라 P_{il} 을 해석하고, 영향력이 큰 독립변수를 선택한다. 즉, 이상 진단의 지표로 P_{il} 값을 계산하여 데이터 타입에 따른 해석 방법을 결정한다. 이 해석 방법에 따라 가장 큰 영향력을 미친 독립 변수를 선택한다.

[0106] MLP(Multilayer Perceptron) 모델 적용 결과, P_{il} 지표를 통해 군집 별 영향력이 높은 변수를 정량적으로 판단할 수 있다.

[0107] 도 34는 MLP 모델 적용 결과의 예를 나타낸다.

[0108] 본 발명의 실시예에 따르면, 설비 이상 진단을 자동화하여 원인이 되는 변수를 신속히 파악함으로써 MTTR(평균 수리시간)을 감소시킬 수 있다. 또한, 설비 이상 패턴을 분류하고 예측함으로써 고장 전 사전 조치를 통해 MTBF(평균 고장 간격)을 증가시킬 수 있다. 또한, 설비의 이상이 발생하기 이전에 사전 조치를 함으로써 제품의 불량을 감소시킬 수 있다.

[0109] 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등)와 같은 저장매체를 포함한다.

[0110] 본 발명의 실시예들은 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 실시예는 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수

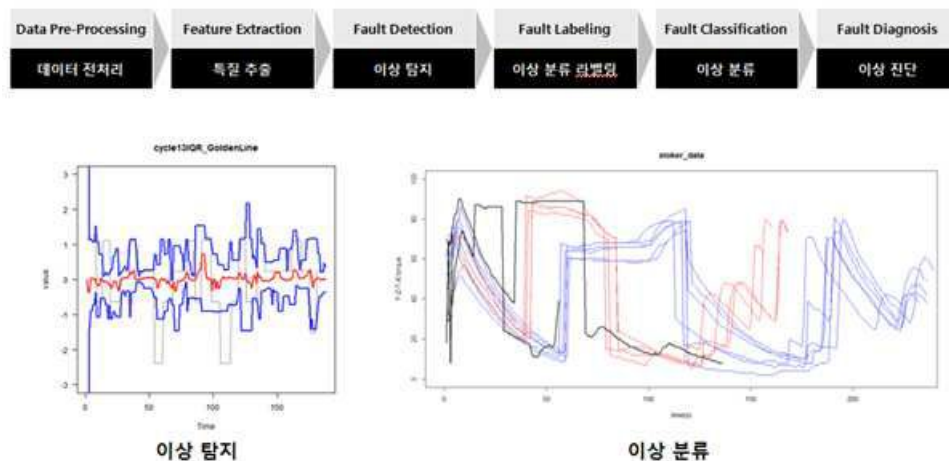
있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 집적 회로 구성들을 채용할 수 있다. 본 발명에의 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 실시예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 실시예는 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. "매커니즘", "요소", "수단", "구성"과 같은 용어는 넓게 사용될 수 있으며, 기계적이고 물리적인 구성들로서 한정되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.

[0111] 실시예에서 설명하는 특정 실행들은 일 실시예들로서, 어떠한 방법으로도 실시 예의 범위를 한정하는 것은 아니다. 명세서의 간결함을 위하여, 종래 전자적인 구성들, 제어 시스템들, 소프트웨어, 상기 시스템들의 다른 기능적인 측면들의 기재는 생략될 수 있다. 또한, 도면에 도시된 구성 요소들 간의 선들의 연결 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것으로서, 실제 장치에서는 대체 가능하거나 추가의 다양한 기능적인 연결, 물리적인 연결, 또는 회로 연결들로서 나타내어질 수 있다. 또한, "필수적인", "중요하게" 등과 같이 구체적인 언급이 없다면 본 발명의 적용을 위하여 반드시 필요한 구성 요소가 아닐 수 있다.

[0112] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

도면1



도면2



도면3

SVNAME
Format
Max Bytes
Value Unit
Min Value
Max Value
Example

It is the name of the status variable.
It is the data format of the equipment constant.
It is the maximum bytes to represent the value of the SV.
It is the unit of the SV.
It is the minimum value of the SV. If the SV is not numeric type, you do not have to fill it.
It is the maximum value of the SV. If the SV is not numeric type, you do not have to fill it.
It is an example of the value.

No	SVNAME	Format	Max Bytes	Value Unit	Min Value	Max Value	Example	Description
SVName001	SV1_USER_CNT	Text	51	Cnt	0	8000000000		Master laser user counter reading Read User counter
SVName002	SV2_USER_CNT	Text	51	Cnt	0	8000000000		Slave user user counter reading Read User counter
SVName003	SV3_T11_CNT	Text	51	Cnt	0	8000000000		Master laser non-reversible total pulse counter reading
SVName004	SV4_T11_CNT	Text	51	Cnt	0	8000000000		Slave laser non-reversible total pulse counter reading
SVName005	SV5_RPT_RATE	Float	4	%	0	650		Master laser repetition rate for manual trigger (depending on speed, current or set value)
SVName006	SV6_PULSE	Float	4	mJ	0	999999		Master laser pulse energy (depending on speed, current or set value)
SVName007	SV7_HIGH_VOLT	Float	4	V	2.3	Data*100		Master laser high voltage (depending on speed, current or set value)
SVName008	SV8_NORM_PSL_CNT	Float	4	Cnt	0	999999		Master laser number of pulses since last New Fill
SVName009	SV9_TUBE_TEMP	Float	4	degC	0	999999	Data*10	Master laser tube temperature
SVName010	SV10_TUBE_PRESS	Float	4	mbars	0	999999	Data*100	Master laser tube pressure
SVName011	SV11_WATER_PRESS	Float	4	mbars	0	999999	Data*100	Master laser cooling water inlet pressure
SVName012	SV12_WATER_TEMP	Float	4	degC	0	999999	Data*100	Master laser cooling water inlet temperature
SVName013	SV13_WATER_FLOW	Float	4	gpm	0	999999	Data*100	Master laser cooling water flow

설비별 계측 변수 추출된 변수 선택된 변수의 시계열 계측데이터

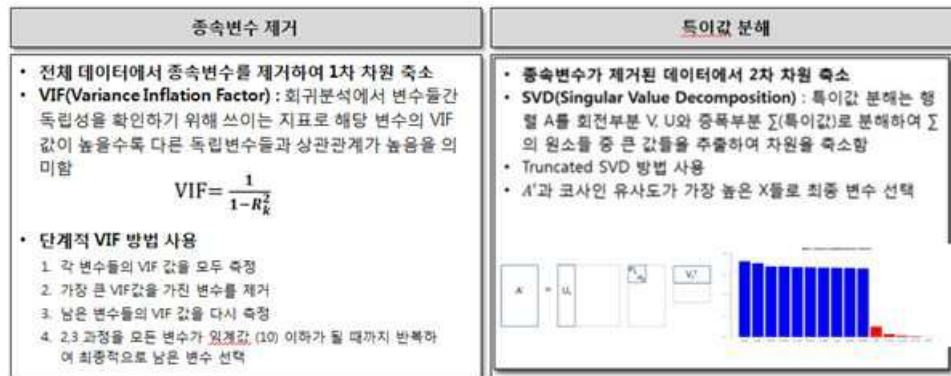
계측데이터의 안정화 구간 설정

도면4

- CVD 설비: 약 300개 센서의 계측 데이터 중 265번 변수의 데이터 포맷

StatusVariableName	EventSaveTime	EventGenerationTime	StatusVariableValue	StatusVariableName	MachineNumber	UnitNumber	Recipe	StepNo	GlauID	LotID	OperID	UnitState
SVName001	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName002	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName003	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName004	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName005	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName006	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName007	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName008	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName009	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName010	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName011	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName012	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName013	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName014	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName015	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName016	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName017	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName018	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName019	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName020	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName021	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName022	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName023	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName024	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName025	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName026	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName027	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName028	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName029	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName030	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName031	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName032	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName033	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName034	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName035	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName036	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName037	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName038	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName039	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName040	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName041	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName042	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName043	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName044	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName045	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName046	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName047	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName048	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName049	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName050	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName051	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName052	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName053	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName054	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName055	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName056	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName057	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName058	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName059	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName060	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName061	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName062	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName063	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName064	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName065	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName066	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName067	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName068	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName069	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName070	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName071	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName072	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName073	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName074	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName075	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H	01058400	01000
SVName076	2020-10-16 10:00:00	2020-10-16 10:00:00	1.0000000000	SV1000000000	4	1	1	1	0	01058400H</		

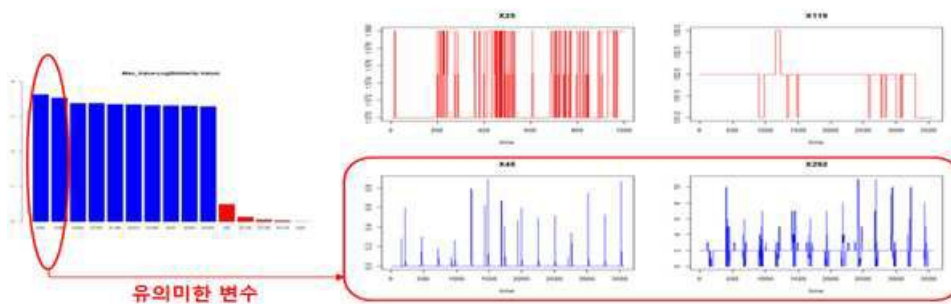
도면5



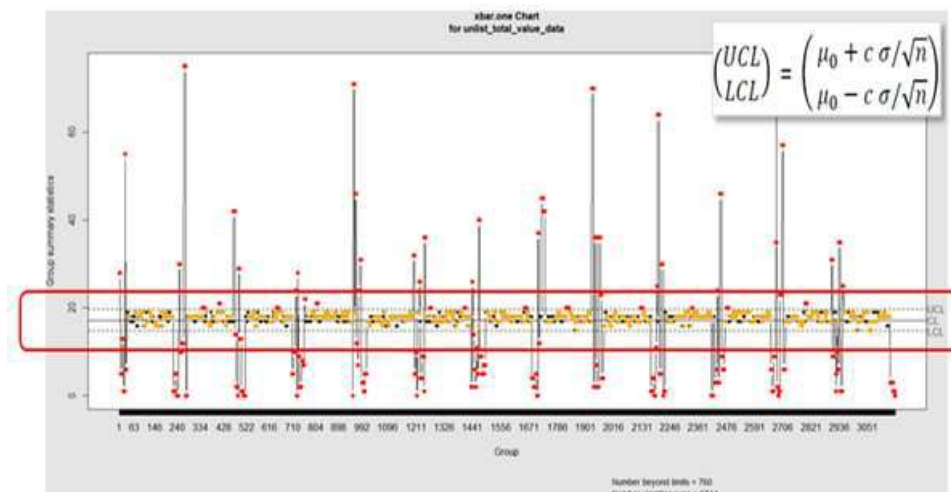
도면6

#6,7,8,19,24,25,37,42,43,44,45,46,47,106,107,118,119,121,122,126,138,
140,141,146,185,255,265,266,292,293,306,319,320,346,347,360,472,47
7,494,498,501,504,524,526,527,559,580,606,608⁴⁾

도면7



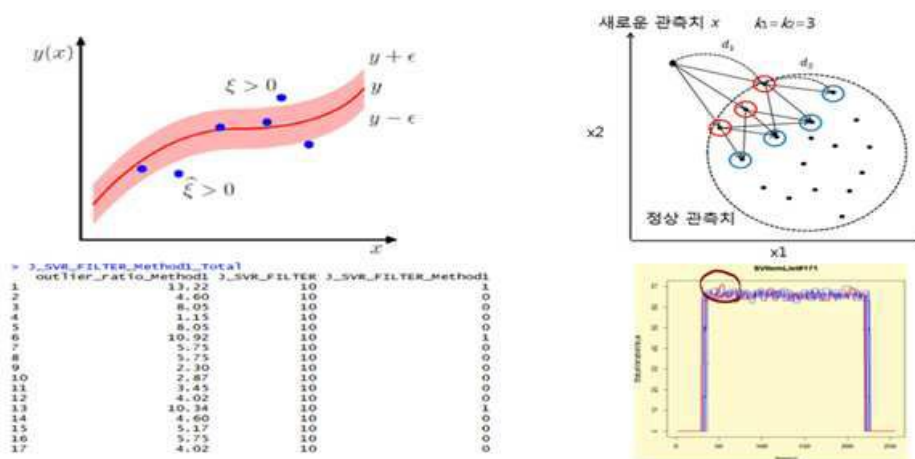
도면8



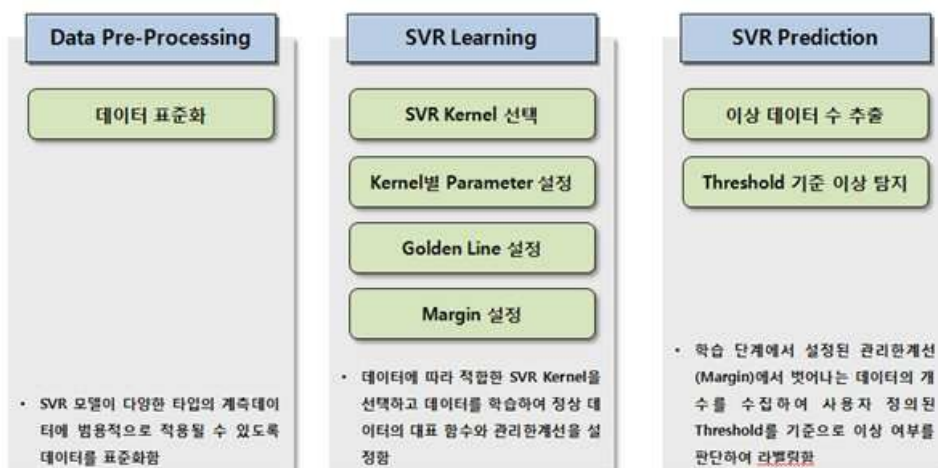
도면9

Output Value	Description
SERIES	한 공정 신호의 주기, EVENT_GENERATION_TIME의 시작과 끝의 차이, 한 시간에 대략 13개 SERIES
UNIT_STATE	공정상태 (JobEventStoredJobReport, JobEventFetchedOutJobReport)
SEQ_SERIES_STEP	Primary Key : Item#번호#Series#Step
CYCLE_TIME	한 공정 신호의 주기, EVENT_GENERATION_TIME의 시작과 끝의 차이
TIME_TO_SATURATION	안정화 상태까지 걸리는 시간 (안정화 상태 : 관리도 안에 들어오는 데이터의 개수가 연속 K개 이상일 경우로 정의)
SATURATION_PERIOD	안정화 상태 구간의 시간
T_MEDIAN	샘플 데이터의 중앙값, 현재 1시간 동안의 데이터를 <u>샘플링하였음</u>
T_UCL	샘플 데이터의 x-bar chart UCL, 현재 1시간 동안의 데이터를 <u>샘플링하였음</u>
T_LCL	샘플 데이터의 x-bar chart LCL, 현재 1시간 동안의 데이터를 <u>샘플링하였음</u>
MEDIAN	Step 별 중앙값
UCL	Step 별 x-bar chart UCL
LCL	Step 별 x-bar chart LCL
MIN	Step 별 최소값
MAX	Step 별 최대값
GRADIENT	Step 별 기울기, 회귀분석 결과에서 추출
SHIFT	Step 별 중심 이동값 (RECIPE에서 생략한 STATUS_VARIABLE_VALUE와의 차이, (Upper Shift : +, Lower Shift : -)
SMOOTHING_VALUE	Double Exponential Smoothing 후 데이터의 평균값 (SPC로 전송하는 대표값으로 활용)
IQR_U	IQR 상한값
IQR_L	IQR 하한값

도면10



도면11

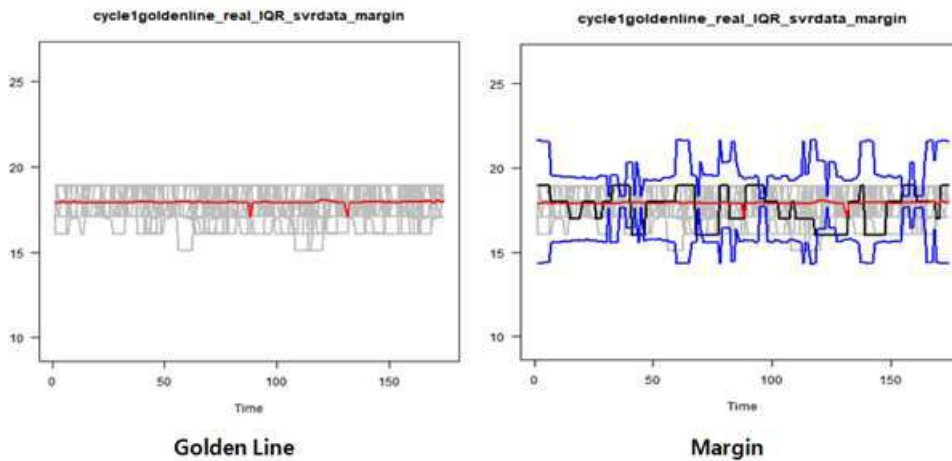


도면12

Kernel	Function	Ksvm(kpar="")
Gaussian Radial Basis Function	$k(\mathbf{x}, \mathbf{x}') = \exp(-\sigma \ \mathbf{x} - \mathbf{x}'\ ^2)$	Sigma
Polynomial	$k(\mathbf{x}, \mathbf{x}') = (\text{scale} \cdot \langle \mathbf{x}, \mathbf{x}' \rangle + \text{offset})^{\text{degree}}$	Scale, offset, degree
Hyperbolic tangent	$k(\mathbf{x}, \mathbf{x}') = \tanh(\text{scale} \cdot \langle \mathbf{x}, \mathbf{x}' \rangle + \text{offset})$	Scale, offset
Laplace Radial Basis Function	$k(\mathbf{x}, \mathbf{x}') = \exp(-\sigma \ \mathbf{x} - \mathbf{x}'\)$	Sigma
ANOVA radial basis	$k(\mathbf{x}, \mathbf{x}') = \left(\sum_{k=1}^n \exp(-\sigma(x^k - x'^k)^2) \right)^d$	Sigma, degree
Linear splines	$k(x, x') = 1 + xx' \min(x, x') - \frac{x + x'}{2} (\min(x, x')^2 + \frac{(\min(x, x')^3)}{3})$	
Bessel	$k(x, x') = \frac{J_{\text{degree}+1}(\sigma \ \mathbf{x} - \mathbf{x}'\)}{\ \mathbf{x} - \mathbf{x}'\ ^{-\text{order}(\text{degree}+1)}}$	Sigma, order, degree

Kernel의 종류

도면13



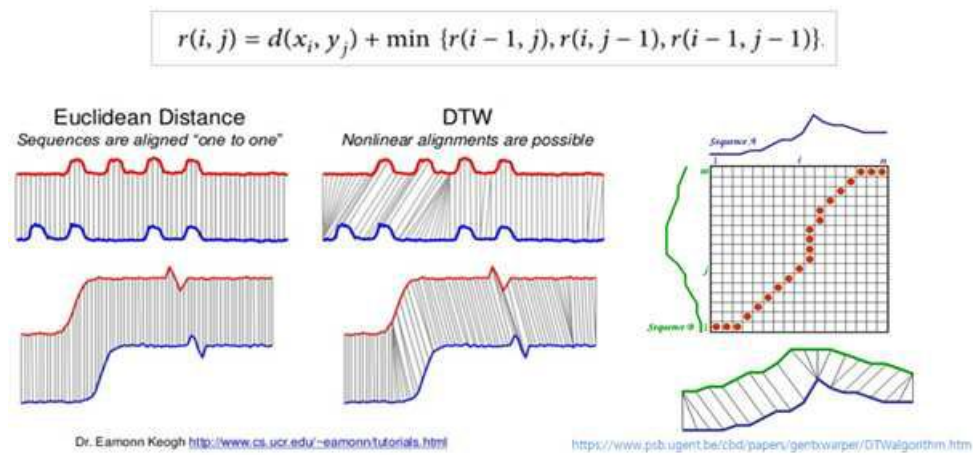
도면14

> J_SVR_FILTER_Method1_Total									
sigma	cost	svrPredictionRMSE	svrPredictionMAE	outlier_num_IQR_svldata_margin	outlier_ratio_Method1	J_SVR_FILTER_Method1	Judge_Method1		
1	7	10	0.1006986	0.09664298	23	13.22	1	abnormal	
2	10	10	0.1010507	0.08063783	8	4.60	0	normal	
3	9	10	0.08083027	0.07902839	14	8.05	0	normal	
4	10	9	0.05543221	0.05088172	2	1.15	0	normal	
5	10	10	0.07446994	0.07336834	14	8.05	0	normal	
6	8	10	0.08952801	0.08778413	19	10.92	1	abnormal	
7	10	10	0.07790128	0.07653378	10	5.75	0	normal	
8	9	10	0.0793674	0.07735515	10	5.75	0	normal	
9	9	10	0.07758106	0.07636224	4	2.30	0	normal	
10	8	10	0.06697159	0.06112194	5	2.87	0	normal	
11	10	9	0.0688603	0.06638845	6	3.45	0	normal	
12	6	10	0.08237007	0.08078943	7	4.02	0	normal	
13	6	10	0.09927904	0.09266682	18	10.34	1	abnormal	
14	6	10	0.0734502	0.07074827	8	4.60	0	normal	
15	8	10	0.08234833	0.07803727	9	5.17	0	normal	
16	7	10	0.06998262	0.06761518	10	5.75	0	normal	
17	7	10	0.06765174	0.06494	7	4.02	0	normal	
18	10	9	0.0609758	0.05855406	8	4.60	0	normal	
19	7	10	0.07123809	0.0688453	7	4.02	0	normal	
20	7	10	0.07145703	0.07010862	2	1.15	0	normal	
21	10	9	0.08701868	0.0854211	13	7.47	0	normal	
22	10	10	0.08444085	0.08390017	16	9.20	0	normal	
23	6	10	0.0698318	0.06335857	13	7.47	0	normal	
24	10	9	0.07070984	0.06958183	8	4.60	0	normal	
25	10	10	0.06810538	0.06722099	4	2.30	0	normal	

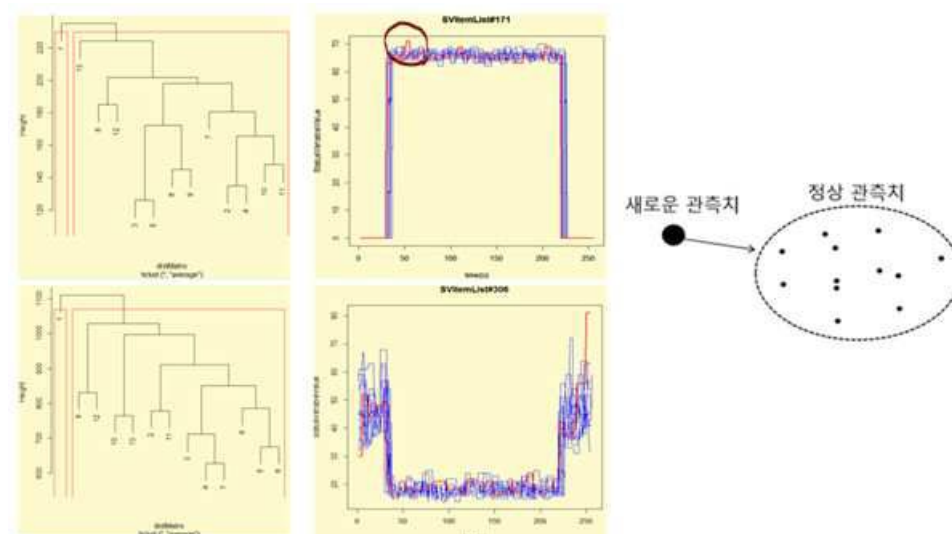
도면15



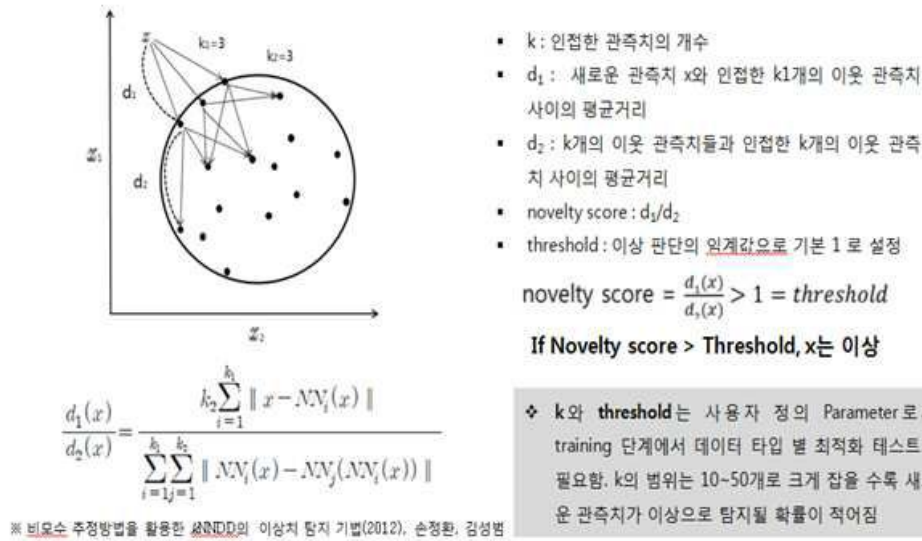
도면16



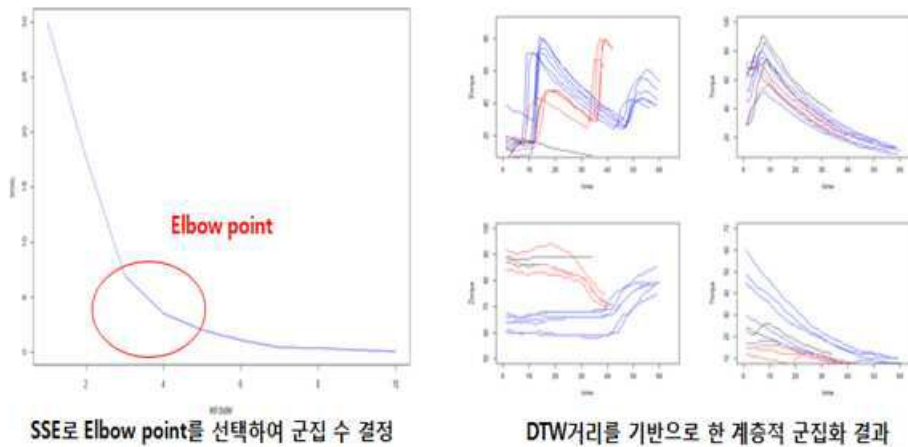
도면17



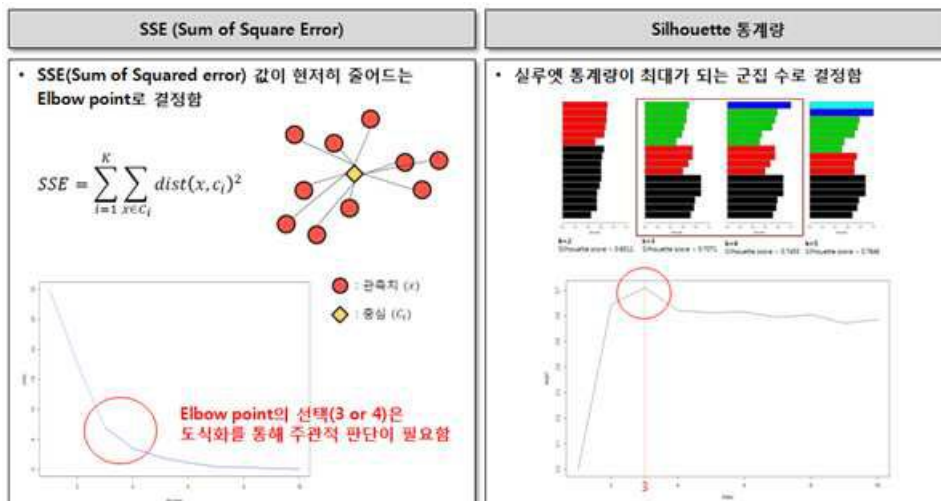
도면18



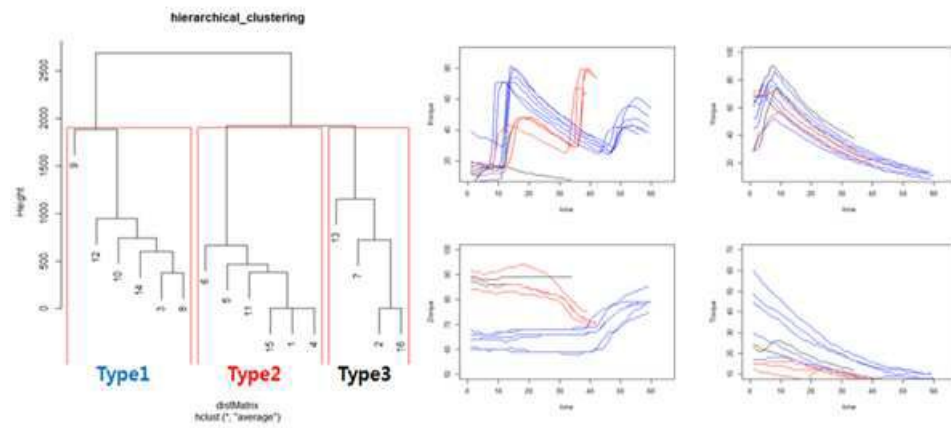
도면19



도면20

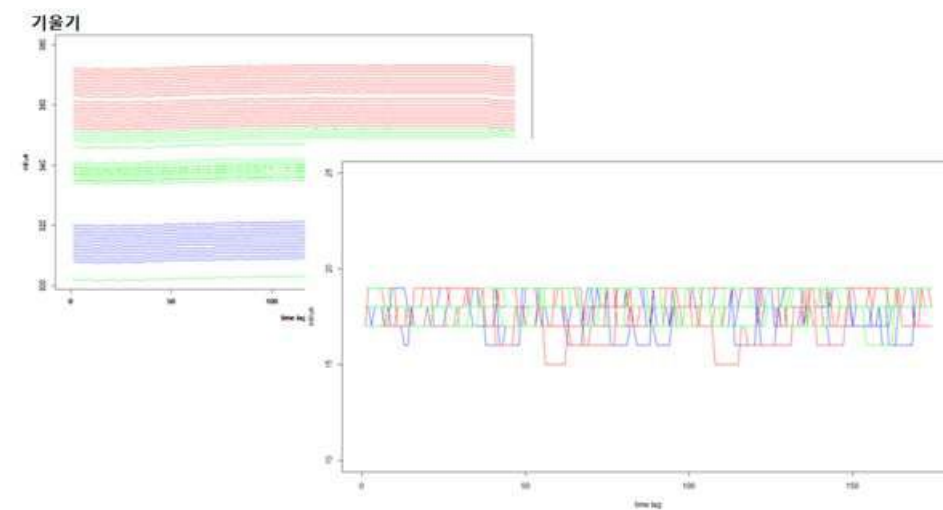


도면21



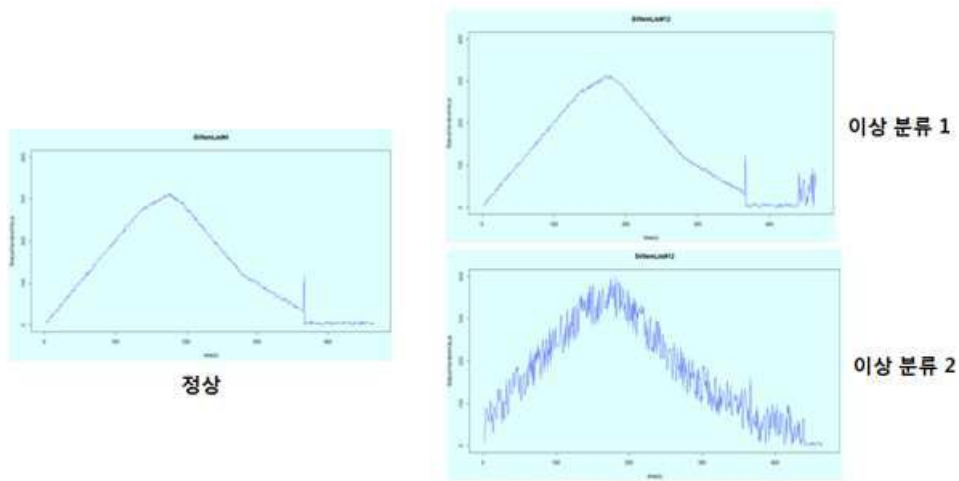
DTW distance를 기반으로 한 계층적 군집화 결과

도면22

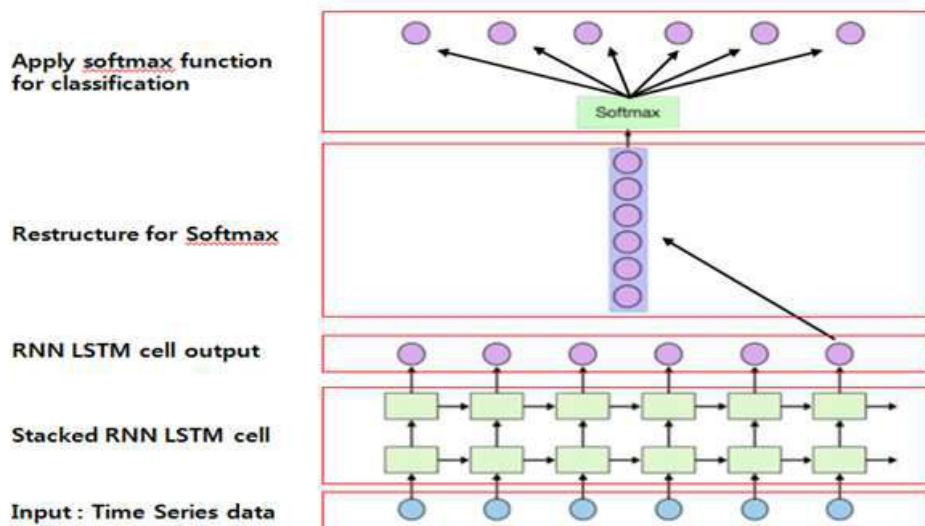


기울기를 대표 값으로 k-means를 수행한 분리형 군집화 결과

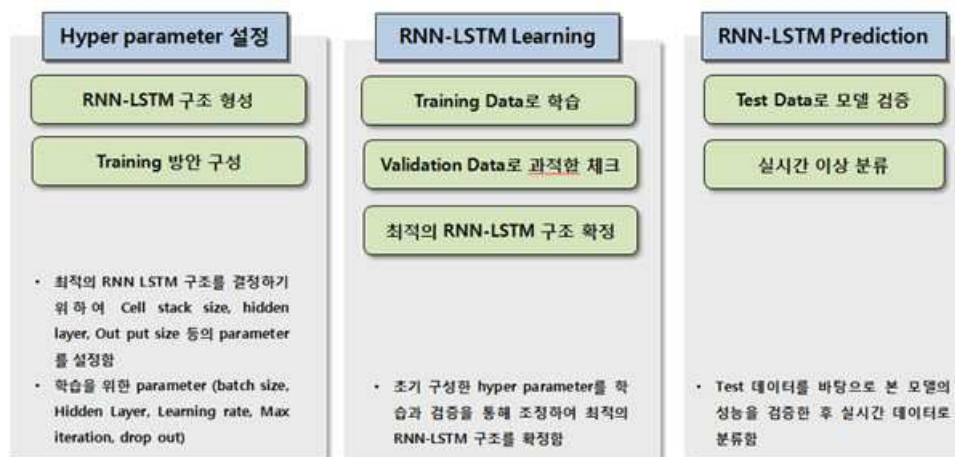
도면26



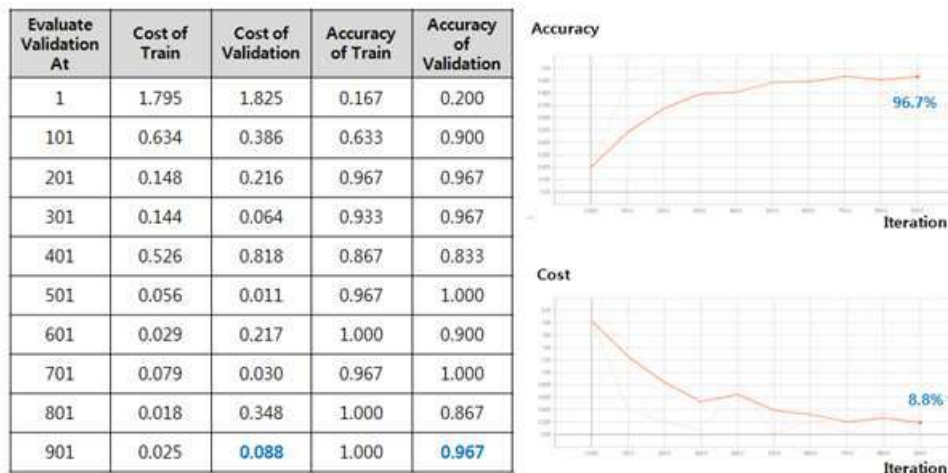
도면27



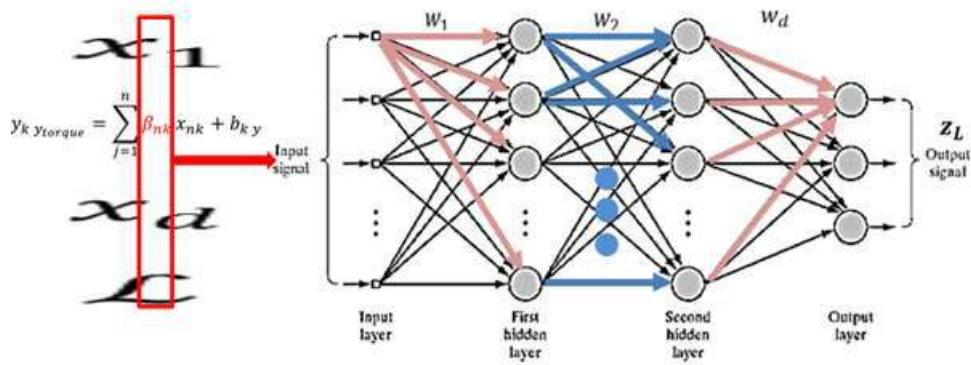
도면28



도면29



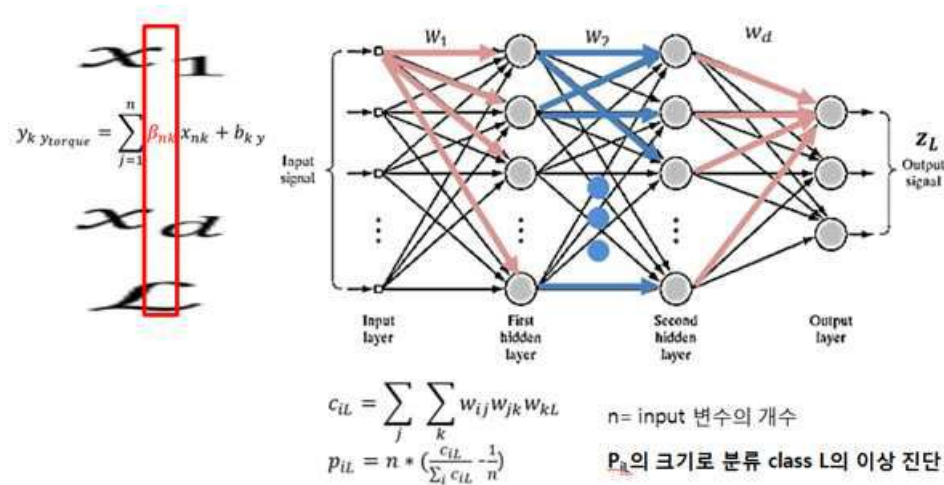
도면30



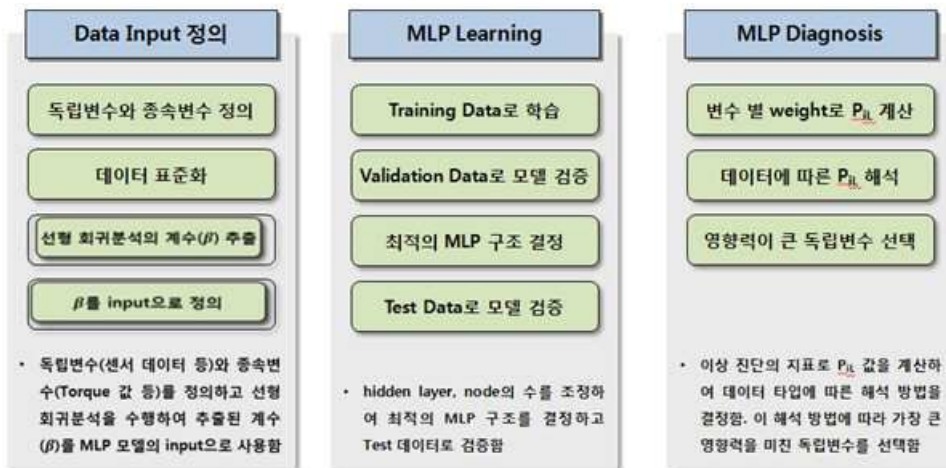
도면31

데이터셋	J_KNDD	J_RNN	J_MLP	J_FINAL (1:1:1)	J_FINAL (1:1:0.5)
Cycle1	1	1	1	1	1
Cycle2	2	2	2	2	2
Cycle3	1	1	1	1	1
Cycle4	1	2	3	1	1
Cycle5	3	2	2	2	2
Cycle6	3	3	3	3	3
cycle7	1	2	1	1	1

도면32



도면33



도면34

군집	Cycle	P_{il}	변수명	Cycle	P_{il}	변수명
1군집	Cycle3	4.0621	Y_OP_LIGHT_LAMP	Cycle10	6.9112	PUT_FORK_STRETCH
		3.9697	HOIST_EQUAL_SPEED		3.9653	TRAVEL_FORWARD
		2.7310	X_PICK_LOAD		2.5815	T_PICK_LOAD
	Cycle8	7.2389	Y_OP_LIGHT_LAMP	Cycle12	4.3528	X_TALIGN_SENSOR2_HALF
		2.7318	TRAVEL_FORWARD		2.7510	HOIST_EQUAL_SPEED
		2.6603	HOIST_EQUAL_SPEED		2.0617	X_TALIGN_SENSOR2_ORIGIN
	Cycle9	3.2396	HOIST_EQUAL_SPEED	Cycle14	5.6014	PUT_FORK_STRETCH
		3.2312	PUT_FORK_STRETCH		3.0209	TRAVEL_PEAK_LOAD_FACTOR1...
		2.8555	TRAVEL_FORWARD		1.8884	Y_OP_LIGHT_LAMP