



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2019년 2월

석사학위 논문

기계학습을 이용한 산업장비의 잔여유효수명 예측

조선대학교 대학원

산업공학과

소민섭

기계학습을 이용한 산업 장비의 잔여유효수명 예측

Remaining Useful Life Prediction of Industrial
Equipment using Machine Learning

2019년 2월 25일

조선대학교 대학원

산업공학과

소민섭

기계학습을 이용한 산업 장비의 잔여유효수명 예측

지도교수 신 종 호

이 논문을 공학 석사학위신청 논문으로 제출함

2018년 10월

조선대학교 대학원

산업공학과

소 민 섭

소민섭의 석사학위논문을 인준함

위원장 조선대학교 교 수 백정우(인)

위 원 조선대학교 교 수 김성준(인)

위 원 조선대학교 교 수 신종호(인)

2018년 11월

조선대학교 대학원

목차

목차	i
그림 목차	iii
표 목차	iv
ABSTRACT	v
제 1 장 서론	1
1.1 연구 배경	1
1.2 연구 개요	3
1.3 논문 구성	4
제 2 장 기계학습	5
2.1 기계학습 소개	5
2.1.1 지도 학습	9
2.1.2 비지도 학습	10
2.1.3 강화 학습	11
제 3 장 유지보수	12
3.1 고장 정의	12
3.2 Prognostics and Health Mangement	13
3.2.1 Prognostics based PoF	14
3.2.2 Prognostics based data	15
제 4 장 잔여유효수명 예측 방법	17

4.1 연구대상	17
4.2 잔여유효수명 예측 방법론	18
4.3 데이터 수집 및 전처리	19
4.3.1 실험 환경 구축 및 데이터 수집	19
4.3.2 평활화 기법을 이용한 노이즈 제거	22
4.3.3 K-means 알고리즘을 이용한 환경 데이터 군집화 ...	27
4.3.4 예지 모델 구현을 위한 데이터 정의	30
4.3.5 학습 모델 정의	32
4.3.6 정규화 및 표준화	34
4.4 예측 모델 평가	37
4.5 Bootstrap 을 이용한 잔여유효수명 예측	47
제 5 장 결론	52
참고문헌	53
감사의 글	60

그림 목차

그림 1. 인공지능, 기계학습, 심층학습의 관계도	5
그림 2. 기계학습 역사	7
그림 3. 기계학습 분류	8
그림 4. 작업 현장의 지도 학습 적용 예	10
그림 5. Hard failure and Soft failure	12
그림 6. Filter	18
그림 7. 분석 수행 절차	19
그림 8. 데이터 측정 센서	20
그림 9. 노이즈 제거 전 시각화	23
그림 10. Simple moving average	24
그림 11. Local regression	25
그림 12. 노이즈 제거 후 시각화	26
그림 13. 군집 시각화 x_2 (1.0 ~ 2.5 미세먼지)	30
그림 14. Support vector machine	34
그림 15. Scatter of real value and predicted value	41
그림 16. 제품 수명 범위	44
그림 17. x_2 (1.0~2.0 미세먼지)의 분포	48

표 목차

표 1. 센서 데이터.....	21
표 2. 데이터 베이스 구조.....	오류! 책갈피가 정의되어 있지 않습니다. 2
표 3. 최적의 군집 개수	오류! 책갈피가 정의되어 있지 않습니다. 8
표 4. x_2 (1.0 ~ 2.5 미세먼지) 군집 결과.....	오류! 책갈피가 정의되어 있지 않습니다. 9
표 5. 하중 (Input) 및 손상(damage) 데이터	32
표 6. (x_1, \dots, x_3) 정규화.....	36
표 7. (x_1, \dots, x_3) 정규화 및 (y)표준화.....	37
표 8. MAPE	39
표 9. 상관 계수.....	40
표 10. T-test	40
표 11. 미수집 운용 데이터에 대한 필터 성능 값 예측	42
표 12. 필터 수명 범위.....	43
표 13. DNN 모델 파라미터.....	오류! 책갈피가 정의되어 있지 않습니다. 6
표 14. Bootstrap을 위한 데이터 집합	50
표 15. Bootstrap을 통해 계산된 필터 평균 수명	51

ABSTRACT

Remaining Useful Life Prediction of Industrial Equipment using Machine Learning

Min-Seop So

Advisor: Prof. Jongho Shin, Ph.D.

Department of Industrial Engineering

Graduate School of Chosun University

In all areas, including industrial field, medical and social, unexpected facility failures can result in unplanned downtime resulting in high cost of loss. In particular, unexpected failures in the industrial field result in a large loss cost associated with equipment and labor. due to these problem. It is necessary to develop technology to prevent unexpected failure by predicting the exact remaining useful life (RUL) of the product and determining the point of facility inspection and replacement beforehand. Accordingly, active research is proceeding recently, such as prognostics and health management (PHM), and condition based maintenance (CBM) method for failure prediction. PHM and CBM are technology that collects operating and environmental data related to the failure/lifetime of the equipment and predict failure that it can

respond in advance before failure. For this purpose, this study proposes a method to predict the exact remaining useful life by attaching sensors to air purifiers to collect operating and environment data, and to calculate the reduction in the performance of equipment according to the use environment. the methodology proposed in this study is signal processing techniques to remove noise signal from sensor data, and clustering, support vector machine, and deep learning, which are machine learning techniques for calculate and predicting reduction rate by use environment.

제 1 장 서론

1.1 연구 배경

최근 제조시스템은 빠르게 발전하는 정보통신기술(Information and Communication Technology, ICT)의 도입으로 4차 산업혁명의 주요 기술 중 하나인 스마트팩토리로 점차 변화하고 있다. 스마트팩토리 환경에서 수집된 새로운 데이터는 제조시스템의 운영을 개선하는데 다양하게 활용될 수 있다. 특히 이를 기반으로 하는 새로운 유지보수 기술의 개발은 점차 증가 추세에 있는 기업의 유지보수 비용을 절감한다는 측면에서 많은 관심을 받고 있다. 미국 산업계의 경우 플랜트 장비 및 기자재의 유지 보수를 위해 사용하는 연간 비용이 2,000억 달러를 넘는다고 하며, 이중에 부적절한 유지보수로 인한 손실 비용이 연간 600억 달러로 전체 유지보수비용의 30%에 이를 것으로 추정하고 있다[1].

유지보수 비용을 절감하기 위해 기업들은 다양한 노력을 해왔으며, 이를 위해 다양한 유지보수기술을 개발하고 적용해 왔다. 일반적으로 사용하는 유지보수 정책은 고장이 발생하는 시점 전후를 기준으로 언제 수행되는지에 따라 사후 정비(corrective maintenance)와 예방정비(prevent maintenance)로 구분할 수 있다[16,17,33]. 사후 정비는 설비의 이상 징후가 감지되었을 때 부품을 교체하거나 수리 하는 등 고장의 사후에 조치를 취하는 방식이다[33]. 이에 반해 예방 정비는 설비의 이상 징후가 감지되기 전에 유지보수 활동을 수행한다[18]. 예방 정비는 다시 시간 기반 정비(time-base maintenance)와 상태 기반 정비(condition-base maintenance)로 나눌 수 있다. 시간 기반 정비는 고장의 예상 발생 시각 이전의 적당한 시점에 주기적으로 유지보수를 수행하는 방법으로 고장이 발생하지 않았을 경우에도 예방 정비를 하는 과잉 정비

가 될 가능성이 존재한다[17,33]. 또한, 시간 기반으로 수행되는 정기적인 유지 보수를 위한 설비 정지로 생산성 저하가 발생하는 단점이 있다. 시간 기반 정비의 가장 큰 문제 중 하나는 다음 유지보수가 수행되기 전에 발생할 수 있는 예기치 못한 설비 고장에 미리 대응할 수가 없다는 점이다.

과거 제조시스템에서 사용되던 단순 설비와 달리 현대 제조 현장에서의 장비들은 기능의 복잡성으로 인해 고장 유형이 다양한 형태를 가지며, 시간과 무관한 고장 유형을 보이는 특징을 가지는 경우가 있다. 따라서 시간 기반 정비에 의한 유지보수는 이러한 제조시스템에서는 설비의 이상에 대응하지 못하게 된다. 이러한 장비의 특성에 따른 유지보수를 수행하기 위해서는 기계의 상태를 실시간으로 파악하여 유지보수를 수행할 수 있는 상태 기반 정비가 필요하다[2].

상태 기반 보전은 다섯 단계의 구성으로 분리 할 수 있다. 첫 번째 설비 운용 중 설비와 관련된 내/외부 데이터를 수집 한다. 두 번째 수집 한 데이터를 특징을 추출하는 신호처리 기법 이다. 세 번째 설비의 상태를 실시간으로 모니터링 하는 진단 기술, 네 번째 설비의 고장 발생 시점을 예측 하는 예지 기술, 다섯 번째 예측한 잔여유효수명 결과를 통해 유지 보수 전략을 수립하는 과정이 있다. 상태 기반 보전을 구성하는 여러 세부 기술 중 예지(prognostics)기술은 최근 활발하게 발전하고 있는 (Internet of Things, IoT) 와 기계학습(machine learning)을 적용하는 방법들이 활발하게 연구되고 있다[51,52]. 예지 기술은 최근 활발한 연구가 시작 된 분야로 발견할 수 있는 방법론이 많은 분야 이다. 따라서 본 연구에서는 센서 네트워크를 통해 실시간으로 수집된 제품 운용 데이터를 활용하여 상태 기반 유지 보수 기술에 필요한 잔여유효수명을 예측하는 방법론을 제안하고자 한다. 제안된 방법론은수집된 운용 환경 데이터를 그룹화 하여 다양한 환경에 따른 제품의 성능 감소 값을 파악하고, 환경의 분포를 파악하여 제품의 잔여유효수명을 예측한다. 이를 위해 사용된 방법론은 Clustering과 Deep Neural Network등의 기계학습 기법과 상관관계 분석 과 같은 다양한 통계 적 기법을 사용한다.

연구에서 제안된 방법론은 공기청정기의 필터에서 수집된 데이터를 사용하여 유용성을 증명하였다.

1.2 연구 개요

본 연구는 설비의 운용을 모니터링할 수 있는 센서를 통해 운용 데이터를 수집할 수 있는 환경에서 제품의 상태를 실시간으로 측정 평가하여 잔존유효수명 및 기대수명을 예측하는 예지 모델 구현을 목표로 한다. 이를 구현하기 위해 설비 운용 데이터 수집과 신호처리 기법을 이용한 노이즈 데이터 제거하여 운용 환경 상태에 따른 군집화하고 기계학습 기법을 적용하여 운용 환경에 따른 잔여유효수명 예측을 수행한다. 본 연구에서 사용되는 세부적인 데이터 수집 및 처리 기법은 다음과 같다.

(1) 센서 네트워크를 이용한 운용 데이터 수집 및 전처리

제품에 센서를 설치하여 운용 데이터를 수집하고, 데이터의 특징을 추출 하기 위하여 Local regression과 Clustering 등 다양한 기법을 사용하여 수집한 데이터를 전처리 한다.

(2) 기계학습을 이용한 운용 환경에 따른 제품 성능 감소 계산

제품의 상태를 실시간으로 확인 할 수 있으며, 수집되지 않은 운용 환경에 대한 제품 성능 감소를 파악하여 제품의 잔여유효수명 예측에 기여 할 수 있는 기계학습 모델을 구현한다.

(3) Bootstrap을 이용한 잔여유효수명 예측

제품의 운용 환경을 고려한 Bootstrap을 사용하여 제품의 잔여유효수명을 예측하는 방법을 소개한다.

1.3 논문 구성

이후 본 논문의 구성은 다음과 같다. 제 2장에서는 제품의 잔여유효수명을 예측하기 위해서 사용된 주요 기법인 기계학습에 대해서 설명한다. 제 3장에서는 유지보수를 위해 제안된 기존의 방법론들에 대해서 논의하고 본 연구에서 제안된 방법론의 장점에 대해서 논의한다. 다음 장에서는 제안된 방법론을 실제 제품 운용 데이터를 사용하여 자세하게 설명한다. 마지막 장은 본 연구의 결론을 간략하게 요약하고 향후 연구 방안에 대해 기술한다.

제 2 장 기계학습

2.1 기계학습 소개

기계학습은 인공지능의 한 분야이며, 인공지능(Artificial Intelligence; AI)은 인간의 지능으로 해결할 수 있는 문제들을 컴퓨터가 모방하는 방법을 연구하는 컴퓨터 공학 및 정보기술의 한 분야이다. <그림 1>은 인공지능과 기계학습 그리고 심층 학습의 관계를 그림으로 표현한 것이다.

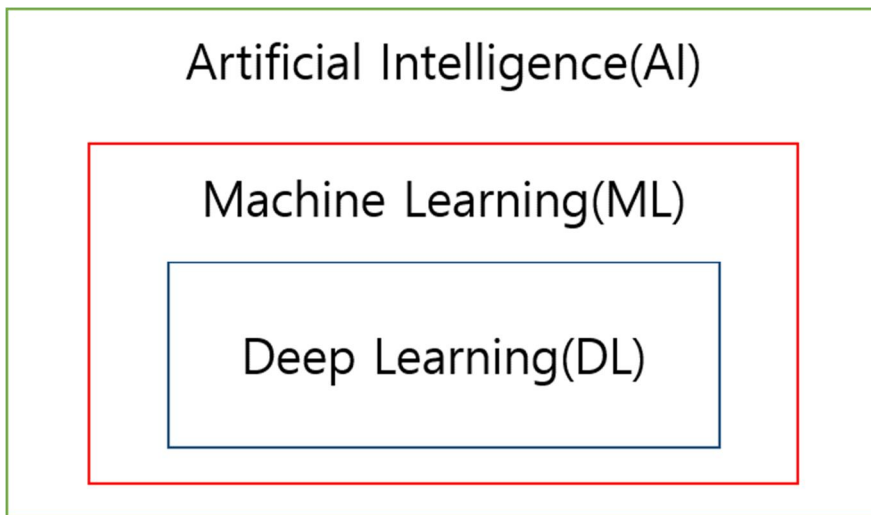


그림 1. 인공지능, 기계학습, 심층 학습의 관계도

기계학습과 심층학습을 포함한 가장 큰 범주인 인공지능은 인간의 지능을 모방하여 인간처럼 행동 할 수 있도록 컴퓨터를 통해 구현 하는 방법이다. 기계학습은 컴퓨터 스스로 수학적 계산을 통해 대용량의 데이터를 분석하여 패턴을 학습하고, 그에 따른 결과를 예측하는 인공지능의 다양한 접근 방법 중 하나이다. 심층 학습은

기계학습의 능력을 한층 더 심화 시키기 위하여 복잡한 계산이 가능한 기계학습을 학습 하는 방법 중 하나이다. 이렇 듯 인공지능과 기계학습 그리고 심층 학습은 모두 사람의 지능을 모방하기 위해 사용한다는 점 에서 밀접한 관련이 있다. 기계학습 이라는 용어의 최초 정의는 “명시적으로 프로그램을 작성하지 않고 컴퓨터에 학습할 수 있는 능력을 부여하기 위한 연구 분야” 로 정의되었다[3]. 초기 인공지능의 대표적인 방법인 전문가 시스템(Expert System)은 인간이 가지고 있는 지식을 데이터베이스화 하여 컴퓨터에 입력한 후 해결하고자 하는 문제에 대하여 인간과 동일한 혹은 그 이상의 문제 해결 능력을 가질 수 있도록 만들어진 시스템이다[53]. 하지만 시대가 발전함에 따라 인간이 알지 못하는 지식을 컴퓨터 스스로 학습하여 해결해 주는 시스템을 요구하기 시작하였고 그 요구를 충족할 방법론이 기계학습이다.

인공지능의 역사적 시작은 Alan Turing의 기계가 생각할 수 있을지를 바탕으로 컴퓨터와 대화를 나누어 컴퓨터의 반응을 인간과 구별할 수 없다면 컴퓨터가 생각할 수 있는 것으로 간주하여야 한다고 주장한 Turing Test에서부터 시작한 것으로 알려져 있다[4]. 한편 Warren McCulloch와 Walter Pitts는 사람의 뇌에서 동작하는 간단한 기능을 흉내 낼 수 있는 최초의 인공 신경망(Artificial neural model)을 이론적으로 증명하였다[5]. 이 논문을 기반으로 1958년 Frank Rosenblatt는 신경망의 기초 알고리즘인 Perceptron을 발표하였다[6]. Perceptron은 출력층의 결과값과 실제 값의 오차가 최소가 되는 가중치를 계산하는 방법이다. 하지만 비선형 문제는 해결하지 못하는 한계점이 있다. 따라서 단층 퍼셉트론의 비선형 문제에 대한 한계점을 해결하기 위해 여러 개의 은닉층(Hidden Layer)으로 구성되어 있는 다층 퍼셉트론(Multilayer Perceptron, MLP)이 제안되었다. 하지만 중간에 은닉층이 없는 다층 퍼셉트론과 달리 MLP의 입력층과 출력층 사이에 존재하는 은닉층의 출력값에 대한 기준값을 정의할 수 없어 학습시키는 방법이 존재하지 않는 문제가 존재하였다[7]. 이는 1986년 Geoffrey Hinton 교수에 의해 다층퍼셉트론을 학습 시킬 수 있는 Back propagation개념으로 해결이 가능하게 되었다[8]. 다층퍼셉트론의 학습 방법 개발은

비선형 문제도 해결할 수 있는 심층 신경 개발의 발판이 되었다. 층수가 많아진 신경망을 딥뉴럴 네트워크(Dep Neural Network, DNN)라고 부르며 이러한 심층 신경망을 학습 시키는 방법을 심층 학습 (Deep Learning, DL)이라 부르며, 이 심층 학습은 기계학습의 대표적인 알고리즘 중 하나이다. < 그림 2 >는 기계학습의 역사를 요약한 그림이다.

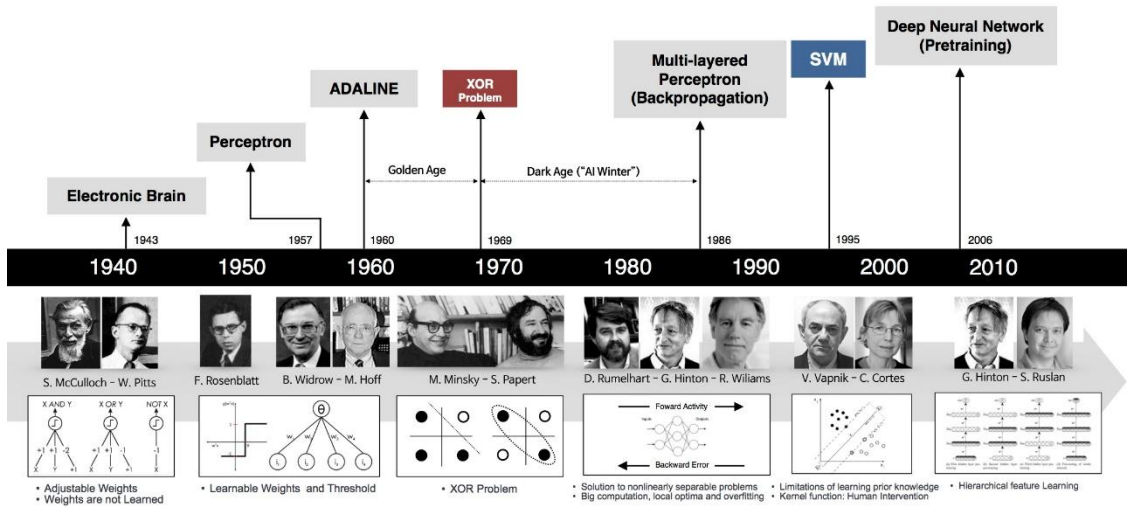


그림 2. 기계학습 역사

[그림 출처: https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html]

1943년 신경 과학자인 Warren McCulloch와 수학자인 Walter Pitts에 의해 처음 인공 신경망을 이론적으로 증명하였다. 1970년대 MLP의 학습 방법에 대한 한계점으로 인공지능은 암흑기로 접어 들게 되었다. 1986년 Geoffrey Hinton교수에 의해 MLP 학습법의 한계 점 에 대한 문제점을 해결 하여 다시 현재까지 활발한 연구가 진행 중이다. 이처럼 인공지능은 약 80년의 역사를 가지고 있으며, 최근 대용량 데이터 수집을 가능케 하는 센서네트워크 기술 및 계산 능력이 향상된 컴퓨터의 발달로 기계학습은 더욱 빠르게 발전할 것이다.

기계학습은 크게 지도학습(Supervised Learning) 및 비지도학습(Unsupervised Learning) 그리고 강화학습(Reinforcement Learning) 세 가지로 구분할 수 있다. 첫 번째 지도학습은 문제에 대한 해답을 컴퓨터가 학습하여 $f(x) \rightarrow y$ 와 같은 함수를 찾아내어 새로운 문제에 대한 답을 찾아내는 미래 지향론 적인 방법이다. 두 번째 비지도학습은 지도학습과 달리 문제에 대한 해답이 존재하지 않는다. 따라서 데이터 간의 유사성을 파악하고, 데이터의 특징을 찾는 방법론이다. 마지막 강화학습은 다양한 환경이 존재하는 상태에서 행동을 취하는 에이전트가 가장 높은 보상을 받기 위해 취해야 하는 행동을 학습하는 방법론이다. 대표적인 예시로 알파고가 존재한다. 본 연구에서는 지도학습과 비지도학습 두 가지 방법론을 사용한다. <그림 3>은 기계학습의 사용 방법에 따라 분류한 그림이다.

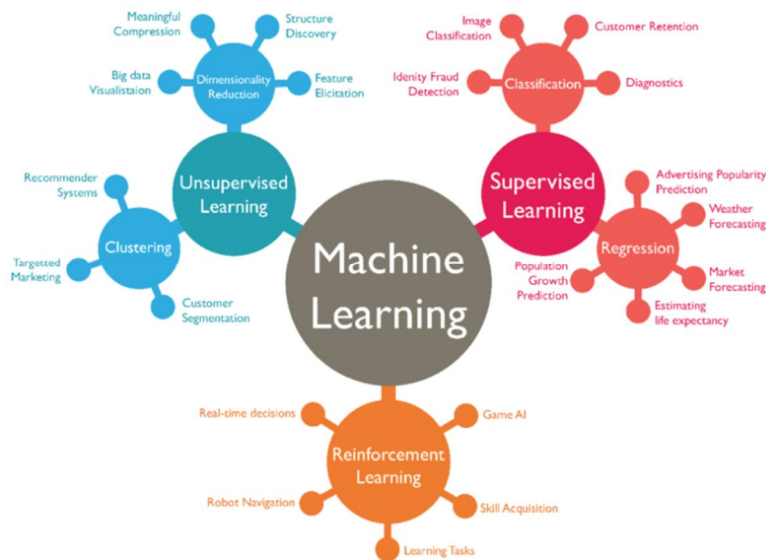


그림 3. 기계학습 분류

[그림 출처: <https://www.slideshare.net/cloudera/advanced-analytics-for-investment-firms-and-machine-learning>]

기계 학습은 목적에 따라 사용 모델을 선택 해야한다. 그 중 지도학습은 임의의 실

수를 예측하기 위한 Regression 모델 및 사진 분류 및 고장 진단 등에 사용할 수 있는 Classification 모델이 대표적인 예이다. 비지도학습은 데이터 집합 중 유사성이 높은 데이터 간 그룹화를 수행하는 Clustering 모델과 영상 처리 및 컴퓨터 비전 등 다차원의 문제를 다루는 곳에서 데이터의 차원을 축소하여 사용하기 위한 모델이 있다. 마지막 강화학습은 게임 및 자율주행차 등 최적의 행동을 찾아 목표를 달성하는 분야에서 사용되는 모델이다.

2.2.1 지도 학습

기계 학습의 방법론 중 하나인 지도 학습은 입력 데이터와 출력 데이터 간의 관계를 하나의 함수로 유추해내어 새로운 입력 데이터에 대한 출력 데이터 값을 예측하는 방법론이다. 지도 학습은 크게 분류(Classification) 및 회귀(Regression) 문제 해결에 적용할 수 있다.

분류 문제는 입력 데이터가 어떠한 클래스(Class)에 속할지 예측하는 방법론이다. 클래스는 입력 데이터에 대응 되는 결과 값으로 두 개의 클래스로 분류하는 경우를 이진 분류라고 하며, 셋 이상의 클래스로 분류하는 경우 다중 분류라고 한다. 이진 분류의 경우 스팸메일에 대한 분류를 한다고 하면, 받은 메일에 대한 출력 값으로 평범한 메일 또는 스팸메일 두가지 중 하나의 값으로 출력된다. 다중 분류의 경우 분류 값이 3개 이상의 출력 값을 가지는 경우를 말한다.

◦ 위험지역 SOP 위반사례 Standard Operating Procedure

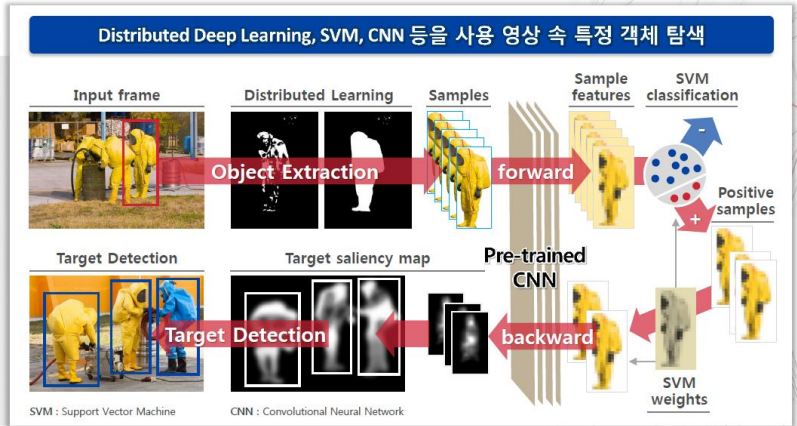


그림 4. 작업 현장의 지도 학습 적용 예[14]

산업현장에서 분류 문제에 적용되는 방법으로는 제품의 양품과 불량을 구분하는 문제[9], 설비의 고장 진단 문제[12] 현장 작업자의 안전장비 착용/미착용 문제[10, 14] 등이 있다.

회귀 문제는 출력 값이 연속적인 숫자를 예측하는 방법으로 예를 들어, 집의 층수, 집 주변의 마트 개수, 지하철까지의 거리 등 다양한 변수들을 활용하여 집값을 예측하는 방법이 회귀 문제에 해당된다. 산업현장에서는 제품의 잔여유효수명을 계산하는 방법[11, 13]이 있다.

2.2.2 비지도 학습

비지도 학습은 지도 학습과 달리 입력 데이터에 대응되는 출력 데이터가 없이 오직 입력 데이터만으로 작업을 수행하는 방법론이다. 대표적인 방법론으로는 군집화(Clustering), Generative adversarial network (GAN)[15]이 있다. Clustering은 입력 데이터들 간의 유사성 동질성 등 비슷한 점을 가진 데이터를 몇 개의 집단으로 분류하는 방법으로 예를 들어 컴퓨터에게 개와 고양이 사진을 입력하여 컴퓨터 스스로 개의 특징과 고양이의 특징을 학습하게 하여 사진을 두 집단으로 분류하게

하는 방법이다. 그리고 최근 비지도 학습의 기법 중 가장 관심을 받는 GAN은 원 데이터가 가지고 있는 확률 분포를 추정하여 비슷한 데이터를 만들어 내는 방법으로 저 해상도 사진을 고해상도로 변환하며, 손상된 이미지를 복원할 수 있다. 또한 최근 논란이 되고 있는 음성 합성, Deep fake 등의 기술이 GAN을 기반으로 한 방법이다.

2.2.3 강화 학습

강화학습은 가장 우수한 보상을 받는 최적의 행동을 취하기 위해 여러 번 문제를 해결 하여 정답을 찾는 방법론이다. 강화 학습에는 행동을 취하는 에이전트, 발생 할 수 있는 모든 상태의 집합과 에이전트가 취할 수 있는 모든 행동 집합, 마지막으로 에이전트의 행동 결과에 따른 보상 값이 필요하다. 강화학습은 세 가지의 데이터 집합을 통해 문제에 맞는 최적의 정책을 찾아내는 방법론이다. 바둑으로 인간을 이긴 알파고와 최근 활발하게 연구되는 자율주행차의 기반이 되는 알고리즘이 기계학습 종류 중 하나인 강화학습이다. 세 종류의 기계학습의 공통된 목표 중 하나는 데이터를 기반으로 미래를 예측하여 의사결정 지원에 도움을 주는 시스템을 만들어 활용하는 것이다.

제 3 장 유지 보수

3.1 고장 정의

고장의 사전적 의미는 시스템, 설비, 혹은 부품이 규정된 기능을 상실하여 동작이 정지하는 것을 말한다. 고장은 크게 hard failure와 soft failure로 나눌 수 있다 [40]. hard failure는 모든 기능이 작동하지 않은 기능 정지형 상태를 말한다[45]. Hard failure가 발생한 경우에는 작업자가 쉽게 고장을 감지하여 바로 대응하는 것이 가능하다[45]. soft failure의 경우에는 부분적인 기능 저하가 발생하는 것으로 완전한 동작의 정지로 이르지 않는지만, 제품에 성능 저하나 순간 정지 같은 문제로 손실이 발생하기도 한다[45]. 이 경우 작업자가 고장을 쉽게 감지할 수가 없어서 hard failure에 비해 고장 진단이 어렵다는 문제가 있다. 최근 많은 연구가 soft failure를 빨리 감지하여 hard failure에 이르기 전에 적절한 유지 보수를 수행하도록 진행되고 있다. < 그림 5 >

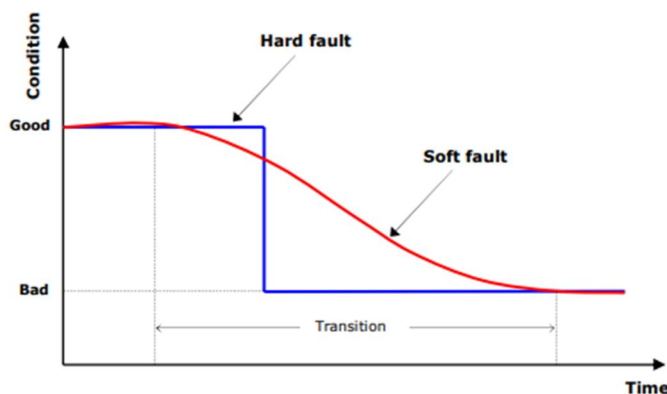


그림 5. Hard failure and Soft failure [45]

장비의 상태를 정확하게 측정하여 soft failure의 진행상황을 인식하고 고장을 예측하는 것은 어렵기 때문에 장비 상태를 수치적으로 계산하여 고장 발생을 예측하는 방법이 필요하다.

3.2 Prognostics and Health Management (PHM)

초기 산업 현장의 경우에 기계 부품의 신뢰도를 확보하기 위해서 사용한 방법으로 제품의 설계단계에서 설계 인자에 대한 불확실성을 정량적으로 고려할 수 있는 신뢰성 기반의 최적설계(Reliability Based Optimization Design)기술이 주목을 받았다 [46]. 그러나 이 방법은 제품이 실제로 운용되면서 발생하는 모든 불확실한 상황을 통제할 수 있도록 설계하는 것이 불가능하다는 한계가 있었다[29]. 이를 보완하기 위해 산업현장에서는 주기적으로 유지보수를 시행하는 예방정비(preventive maintenance, PM)를 기반으로 유지보수를 수행하였다. 그러나 이는 서론에서 설명한 것과 같은 문제점을 가지고 있다. 최근 이러한 문제를 해결하기 위해 고장 예지 및 건전성 관리를 접목한 상태 기반 정비 기술에 대한 연구가 활발히 진행되고 있다. 고장 예지 및 건전성 관리 기술(PHM)은 크게 설비의 이상 징후를 모니터링 하여 진단하는 진단기술(diagnostics)과 잔여유효수명(Remaining Useful Lifetime, RUL)을 예측하여 설비를 효율적으로 관리하는 예지 기술(prognostics)로 나누어진다[31]. 앞의 진단기술은 설비의 성능이 이미 한계에 도달한 후에 적용된다는 점에서 정비를 계획하고 수행하는 부분으로 사용하기에는 한계가 있다[32]. 이에 반해 예지 기술은 설비의 상태에 대한 평가를 기반으로 잔여유효수명을 예측하여 최적의 유지보수 정책을 수행하도록 한다 [33]. 따라서, 상태 기반 유지보수에서 예지 기술은 가장 핵심적인 기술 중의 하나이

다. 예지 기술은 크게 물리모델기반방법(PoF based degradation model)과 데이터기반 방법(Data-driven approach)인 두가지로 나눌 수 있다[33,34]. J.z. sikorska [36]는 예측모델을 선택하기 위한 여러 모델들의 이점 및 단점에 대하여 설명하였다. 전문가 시스템의 경우 이해하기 쉬우며 개발이 간단하다는 장점이 있는 반면 인간이 가지고 있는 지식의 한계점에 대한 문제가 있으며 정확한 잔여유효수명을 예측하는 것은 불가능하다고 하였다. 또한 ARMA(Auto regression moving average) 모델의 경우 세부 고장 메커니즘에 대한 사전 지식이 필요하지 않으며 단기간 예측에 강한 반면 장기 예측의 경우 신뢰도가 낮다고 하였다. ANN(Artificial neural network)을 사용할 경우 복잡한 다차원의 비선형 시스템을 컴퓨터 소프트웨어를 사용하여 모델링 할 수 있는 반면 많은 데이터를 필요로 하며 최근 연구중인 분야로 아직 신뢰할 수 없는 부분이 있다. 마지막 물리적 모델의 경우 예측의 정확도가 가장 높으며 출력물 결과를 쉽게 이해할 수 있지만 상세하고 완벽한 물리적 모델이 필요하며 모델이 개발 된 실험 조건에 영향을 받아 다양한 환경에서의 예측 또한 고려할 필요가 있다.

3.2.1 Prognostics based PoF (Physical of Failure)

물리모델기반 방법은 설비에 가해지는 다양한 고장메커니즘에 대한 지식을 활용하여 제품의 상태를 수치적으로 평가하는 방법이다[28,34]. 고장메커니즘은 두개의 범주로 나누어진다[19]. 첫번째 제품에 가해지는 응력이 제품이 가진 강도를 초과하여 고장이 나는 경우이다. 두번째는 제품의 피로에 의한 고장으로 마모, 부식, 균열 등 사용기간에 따라 제품에 누적되는 손상에 의한 고장이다. Abhigit [19]는 고장메커니즘에 영향을 미치는 여러 원인들을 설명하였으며, Tiedo tinga [20]는 다양한 물리적 모델을 사용하여 고장을 예측한 사례들을 설명하였다. Jie gu [22]는 인쇄 회로 기판에 가해지는 진동 부하를 Miner`s 규칙[23]을 사용 제품의 잔여유효수명을 계산하는 방법을 제안하였다. Jing qiu [21]는 베어링의 진동 분석과 손상 역학에 기반하여 잔

여유효수명을 예측하는 방법을 제안하였다. Oppenheimer [24]은 균열 성장 법칙을 기반으로 rotor shaft의 잔여유효수명을 추정하기 위해 물리적 모델 방법을 사용하였다. G.J. Kacprzynski[25]는 헬리콥터 기어의 잔여유효수명을 예측하기 위한 물리적 모델을 제안하였다. Lijie chen[26]은 지수 함수에 기반한 MLDS (Modified linear damage summation) 방법을 사용하여 터빈의 피로 수명을 예측하였다. Gong[38]은 해양플랜트의 무어링의 피로 수명을 예측하기 위하여 S-N curve, Miner`s rule을 적용하였다.

PoF 기반 모델은 다른 모델들보다 우수한 정확도를 가지고 있다[20,34]. 따라서, 발전소, 비행기 등 안전성이 최우선 되는 분야에서 꼭 필요한 방법이다. 그러나 모든 고장에 대하여 물리적 모델을 만드는 방법은 매우 어려워 사용하는데 있어서 제한적이다[27]. 또한 제품의 수명 예측은 전체 시스템이 아닌 시스템의 구성품에 대한 개별적 예측을 수행하고 있다. 따라서 전체 시스템의 건전성을 예측하기 위해 사용하는 것에 대한 한계가 있다.

3.2.2 Prognostics based data

4차 산업혁명의 핵심 구성요소 중 하나인 센서네트워크 기술 개발은 현장에서 운용 중인 제품에서 실시간으로 다양한 데이터를 원격으로 수집하는 것을 가능하게 하였다. 그로 인해 최근 데이터를 기반한 유지보수 정책이 활발하게 연구되고 있다. 데이터기반방법의 하나의 예는 설비에 가해지는 하중(Input)대비 손상(damage) 데이터를 수집, 두 데이터간의 관계를 학습하여 제품 운용 중 받는 하중에 대한 손상을 계산하여 잔여유효수명을 예측하는 방법론이다. Xiao-shenh si[35]는 통계적 데이터를 기반으로 예지 모델을 구현하여 잔여유효수명을 추정한 다양한 연구를 리뷰하였으며, Kamal medjaher [33]은 MoG-HMM (Mixtyre of Gaussians Hidden Markov Model) 을 사용하여 베어링의 잔여유효수명을 계산 하였다. M-Xie[37]는 Autoregressive integrated moving average (ARIMA)와 Recurrent neural network (RNN) 모델을 사용하여 압축기의 고장을

예측 하였다. 하지만 고장을 예측 하는 성능이 단기간 예측에 높은 성능을 보이는 반면 장기간 예측에서는 좋지못한 성능을 보인다는 단점이 존재한다. Chao hu[39]는 리튬이온 배터리의 충전 상태와 용량 상태를 추정하기 위해 Extended Kalman Filter를 기반으로 한 MultiScale Framework를 사용하였다.

데이터 기반 예지는 최근 다양한 운용 데이터가 수집 가능한 현 상황에 가장 활발히 연구 중인 방법론이다. 본 연구에서는 운용 환경에 따른 필터에 성능 감소값을 계산하며, 두 변수 간의 상호 관계를 학습하여 수집되지 않은 환경에 대한 필터 성능 감소 값을 예측한다. 그 후 필터의 운용 환경의 분포를 파악 제품의 잔존유효수명을 예측하는 데이터 모델을 기반으로 한 PHM 개선을 목표로 한다.

제 4 장 잔여유효수명 예측 방법

4.1 연구대상

본 연구에서 제안될 방법론은 공기청정기 부품 중 하나인 필터를 대상으로 개발되었다. 공기청정기에 설치된 필터는 공기 중에 포함된 여러 가지 오염물을 제거하는데 사용되는 부품으로 사용 시간이 증가함에 따라 성능이 저하되어 적절한 시점에 교환이 필요한 소모품이다. 오염물을 제거하는 필터는 기능적 범용성으로 인해 가정·자동차·발전소 등 여러 가지 제품에서 다양한 형태로 사용된다. 본 연구에 사용된 공기청정기는 프리필터와 탈취필터 및 마이크로 집진필터로 구성된 3종류의 필터를 사용하고 있다. 첫 번째 프리필터는 공기 중에 큰 입자를 가진 먼지를 제거하는 목적으로 설치되며 반 영구적으로 사용이 가능하다. 두 번째 탈취 필터는 암모니아나 톨루엔 등과 같은 여러 유해가스 및 생활 악취를 제거 하는 필터로 6개월에서 12개월을 주기로 교체를 권장하고 있다. 마지막으로 본 연구의 대상인 마이크로 집진필터는 프리필터에서 제거 할 수 없는 오염물을 정전기를 이용하여 필터로 흡착하여 공기중에서 제거하는 방식을 사용한다. 마이크로 집진필터는 탈취 필터와 같이 일정한 주기의 교체가 필요하다.

일반적인 필터의 교체 시점 결정은 설비가 운용된 일정 시간마다 이를 교체하는 방법을 사용한다. 이 방법은 서로 다른 사용 환경에 따른 필터의 성능 감소 변화를 고려하지 않기 때문에 필터의 남아있는 사용 수명에 관계없이 교체가 발생하는 문제가 있다. 이를 개선하기 위해 필터에 포집된 먼지에 의해 발생하는 압력 변화를 실시간으로 측정하여 필터의 성능을 측정하는 방법이 사용되기도 한다. 이 방법은 제품의 실시간 상태를 평가할 수 있는 장점이 있는 반면에, 제품의 상태 예측에 따른 정비를

할 수 없다는 점에서 한계가 있다. 따라서, 본 연구에서는 필터의 상태 측정과 예측을 바탕으로 정확한 교체 시점을 결정할 수 있는 방법론을 제안하여 보다 효율적인 유지보수를 지원하도록 하는 것을 목표로 한다.

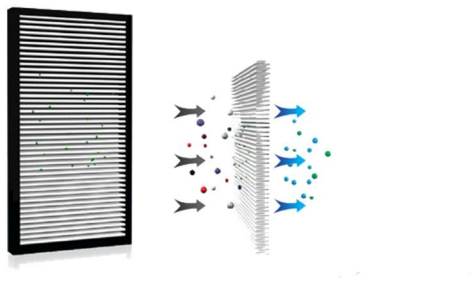


그림 6. Filter (source: <http://kor.pngtree.com>)

4.2 잔여유효수명 예측 방법론

필터의 상태 예측은 잔여유효수명으로 계산이 되고, 잔여유효수명은 외부 사용 환경에 따른 필터의 성능 저하 정도와 관련이 깊다. 이러한 점을 반영하여 운용 환경에 따른 제품의 잔여유효수명을 예측하는 방법론을 개발하였고, 본 연구에서 제안된 방법은 크게 네 단계로 구성되어 있다. 첫 번째 단계는 잔여유효수명 예측 모델에 사용할 운용 데이터 수집 및 수집한 데이터의 전처리로 구성되어 있다. 두 번째 단계는 운용 환경에 따른 수명의 감소를 예측하기 위한 기계 학습 모델을 설계하고 수집된 데이터로 모델을 학습한다. 세 번째에서 학습된 모델을 사용하여 운용 환경에 따른 성능 감소를 예측한다. 네 번째 Bootstrap을 통해 제품 잔여유효수명을 예측하는 과정으로 이루어져 있다.

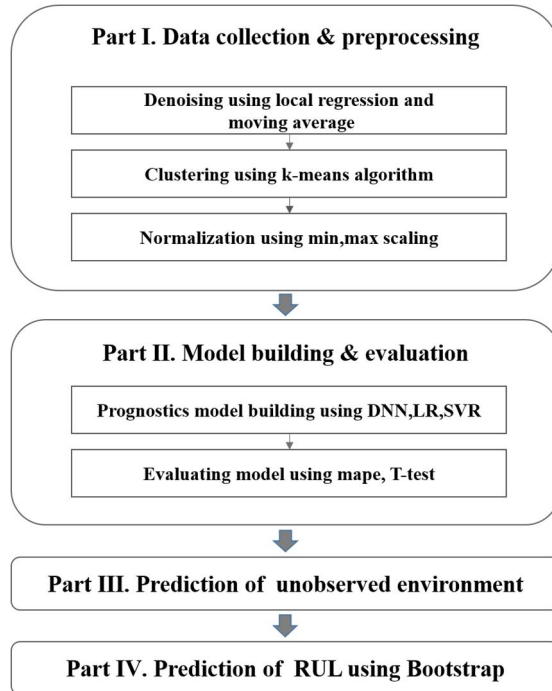


그림 7. 분석 수행 절차

4.3 데이터 수집 및 전처리

4.3.1 실험 환경 구축 및 데이터 수집

센서를 통해 수집하는 데이터는 크게 모니터링하는 제품의 상태를 나타내는 성능 지표와 제품이 사용되고 있는 환경을 기록한 환경데이터로 구분할 수 있다. 본 연구의 대상인 집진필터는 오염물이 필터에 쌓이면 성능이 저하되는 특성이 있다[41]. 필터에 포집된 오염물질이 많아지면서 공기 흐름의 저항이 증가하고 이는 필터를 통과하는 공기의 유량 감소로 나타난다. 이러한 특성을 이용하여 필터의 성능 지표로 필터를 통과한 공기의 유량을 측정하여 사용한다.

환경데이터는 필터의 성능저하에 영향을 미치는 외부 환경을 의미하며, 본 연구에

서는 크기별 미세먼지와 온/습도 세 가지 종류의 환경데이터를 수집한다. 필터의 오염에 따른 성능 저하는 공기 중 미세먼지의 양에 직접적으로 영향을 받기 때문에 이를 측정한다. 미세먼지의 크기에 따른 성능 저하 영향이 차이가 발생할 수 있기 때문에 미세먼지는 센서가 측정할 수 있는 크기인 가장 큰 입자 2.5 ~ 10 (um), 중간 입자 1.0 ~ 2.5 (um), 가장 작은 입자인 0.3 ~ 1.0(um) 세 종류로 구분하여 측정한다. 또한 미세먼지가 필터에 흡착되는데 필요한 정전기가 온/습도에 영향을 받기 때문에 두가지 값도 환경 데이터로 함께 수집한다[42]. < 그림 8 >과 < 표 1 >은 공기청정기에 부착되어있는 센서의 위치와 단위를 나타낸다.

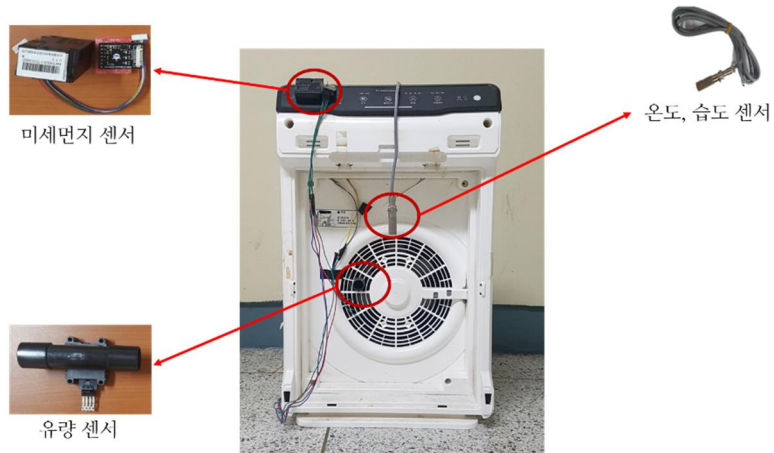


그림 8. 데이터 측정 센서

표 1. 센서 데이터

Variable	Data type	unit
x_1	0.3 ~ 1.0 미세먼지	ug
x_2	1.0 ~ 2.5 미세먼지	ug
x_3	2.5 ~ 10 미세먼지	ug
x_4	온도	°C
x_5	습도	%
y	유량	SLPM

분석에 사용된 데이터는 35일간 수집된 운용 데이터이며, 여섯 종류의 데이터가 세 종류의 센서에서 수집되었다. 첫 번째 센서는 크기에 따른 공기 중 미세먼지(x_1, \dots, x_3)의 양이 ug(Microgram) 단위로 측정되며, 세 종류의 미세먼지 데이터는 하나의 미세먼지 센서에서 동시에 수집한다. 두 번째 온도(x_4) 및 습도(x_5) 데이터는 온/습도를 동시에 측정할 수 있는 센서로 수집하며 마지막 필터의 성능 측정 지표인 유량(y) 데이터는 유량 센서를 통해 SLPM(Standard liters per minute) 단위로 수집된다. 공기청정기에 설치된 센서는 Arduino를 통해 전류를 공급 받게 되어 데이터를 측정 한다. 측정된 데이터를 저장 하기 위해 컴퓨터의 LabVIEW 소프트웨어를 통해 Arduino를 제어하며, 센서에서 측정되어 출력되는 데이터는 Txt파일 형태로 구성된 데이터 베이스에 저장된다. 제품을 실제 35일간 사용하면서 각 2초 간격으로 수집된 데이터가 데이터 베이스에 저장이 되며, 한번 기록시 모든 센서의 측정값이 동시에 기록이 된다.

< 표 2 >는 데이터베이스에 저장되어 있는 센서 데이터들의 구조를 보여주고 있다.

표 2. 데이터 베이스 구조

관측 번호	$x_1(0.3 \sim 1.0$ 미세먼 지)	$x_2(1.0 \sim 2.5$ 미세먼 지)	$x_3(2.5 \sim 10$ 미세먼 지)	x_4 (온도)	x_5 (습도)	Y (유량)
1	29	29	29	19.94	58.72	22.91
2	29	29	29	19.94	58.72	23.31
3	29	29	29	19.93	58.72	24.35
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1499998	24	35	39	18.02	40.32	21.15
1499999	27	35	41	18.00	40.32	22.11
1500000	27	35	41	17.99	40.29	20.60

데이터 베이스의 첫 번째 열은 수집된 150만 개의 데이터를 시간 순으로 관측 번호를 부여한 것이고, 두 번째 열부터 다섯 번째 열까지는 관측 시점에서의 환경 데이터가 기록된다. 여섯 번째 열은 관측된 환경에서 성능 지표인 유량의 값이 기록된다.

4.3.2 평활화 기법을 이용한 노이즈 제거

센서를 통해 수집된 데이터를 효과적으로 활용하기 위해서는 전처리 과정을 거쳐 데이터의 특징이 담겨 있는 의미 있는 데이터를 추출해야 한다. 본 절에서는 데이터 전처리 과정 중 데이터의 노이즈를 제거하기 위한 방법을 설명한다. 노이즈는 측정된 데이터에 무작위 또는 통계적으로 잡음이 포함되는 것을 말하며, 과도한 노이즈는 데이터의 분석 결과에 의미 없는 결과값을 도출하게 한다. < 그림 11 > 은 센서 데이터 노이즈를 제거 하기 전 $x_1(0.3 \sim 1.0$ 미세먼지)와 y (유량) 데이터를 시각화 한 것이다.

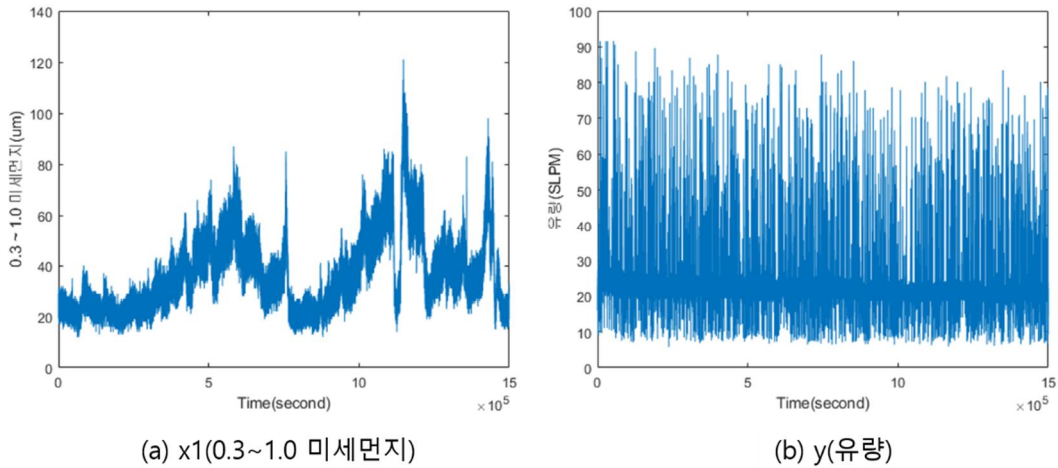


그림 9. 노이즈 제거 전 시각화

온도 및 습도 등 환경은 급변하지 않고 일정 시간을 두고 천천히 변화한다. 하지만 본 연구에서는 수집된 센서 데이터 x_1 (0.3 ~ 1.0 미세먼지)의 경우 초 단위로 진폭이 급격히 존재하는 것을 확인할 수 있다. 이는 센서의 측정값에 대한 신뢰도가 높지 않고 너무 민감하게 반응하는 특성에 따른 결과이다. 이러한 급격한 측정값을 변화는 상대적으로 변화가 더딘 환경 특성과는 맞지 않기 때문에 측정된 값에 노이즈가 존재했다고 생각 할 수 있다. y (유량) 데이터의 경우 필터에 시각화를 통해 확인한 결과 먼지가 포집되어 시간이 지남에 따라 필터를 통과하는 공기의 유량이 감소해야 함에도 불구하고 노이즈로 인해 성능 저하의 뚜렷한 특징이 드러나지 않는 것처럼 보인다. 따라서 본 연구에서는 수집된 센서에 노이즈를 제거하고 실제 값을 분석에 사용하기 위해 다양한 기법을 사용하여 노이즈를 제거한다.

노이즈를 제거하는 방법으로는 kalman Filter, Wavelet Transform, Smoothing spline,

simple moving average, local regression 등 여러 방법이 존재한다. 이중에서 본 연구는 simple moving average 그리고 local regression [50]을 사용하였다. simple moving average는 전체 데이터 중 일정 시점을 기준으로 연속된 일부 데이터 n개를 사용하여 순차적으로 평균을 구하는 방법이다 (그림 10 참조). local regression은 전체 데이터 중 특정 x_0 값 근처의 관측자료를 사용하여 최소자승법을 적용해가는 방법으로 두 기법은 변동성이 존재하는 데이터를 평활화 하여 노이즈를 제거하는 방법이다. local regression의 경우 차수가 3,4차 이상인 경우 특정 x_0 값 근처의 nearest neighbor을 찾기 힘들어 성능이 낮아진다. 따라서 y (유량)의 값은 시간이 지남에 따라 감소하는 성능 지표인 y (유량) 데이터는 시간에 따라 증가 하고 감소하는 환경 데이터에 비하여 낮은 차수로 데이터를 설명 할 수 있으므로 local regression을 사용하였다. 운용 환경데이터인 (x_1, \dots, x_5) 의 측정값들은 simple moving average를 사용하였다.

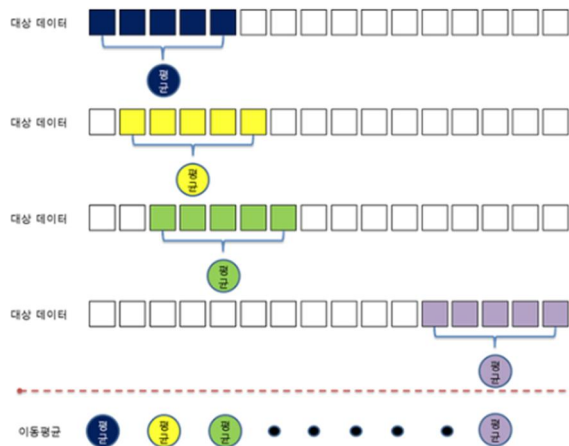


그림 10. Simple moving average

[그림 출처: <http://blog.naver.com/PostView.nhn?blogId=ntkor&logNo=199891958>]

Simple Moving Average의 사용된 식은 아래와 같다.

$$F_t = \frac{x_{t-n} + \dots + x_{t-1} + x_t}{n}$$

F_t 는 t 시점을 기준으로 이전 n 개 데이터의 평균을 계산한 데이터의 평균 값을 의미하며, 마지막 값인 F_n 까지 데이터를 계산한다. 예를 들어 데이터 집합 $X = (2, 3, 7, 4, 6, \dots)$ 이 존재 하고 n 은 3이라고 가정 하면, $\left(\frac{2+3+7}{3}\right) = 4$ 의 값이 F_1 의 값이 되며, $\left(\frac{3+7+4}{3}\right) = 4.6$ 이 F_2 값이 되는 것이다. 계산이 완료 되면 데이터의 짧은 주기에 존재하는 노이즈 데이터가 평균으로 대체되어 노이즈가 제거 된다. 본 연구에 사용한 n 값은 10000개이다.

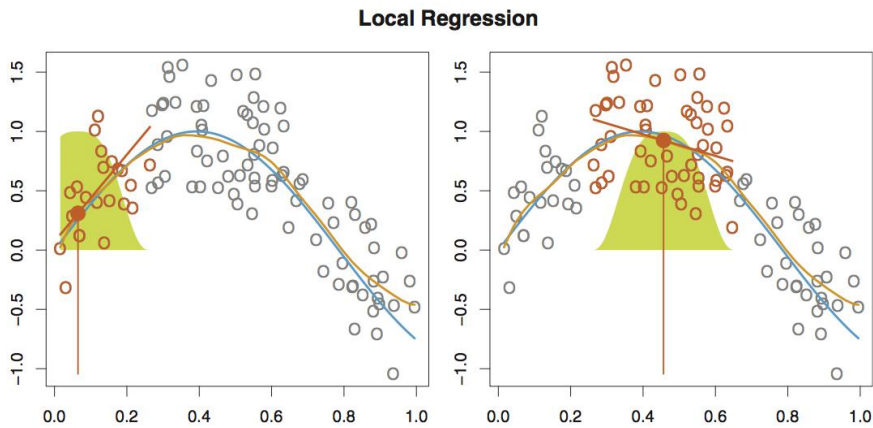


그림 11. Local regression[50]

Local regression의 알고리즘은 네 단계로 구성되어있다[50]. 첫 번째 0~1사이의 값을 가지는 파라미터 s 값을 선택한다. s 는 $x_1 \sim x_n$ 개의 데이터 집합 중 x_l 에서 가까운

일부분의 데이터를 선택하기 위한 파라미터 값이다. 파라미터 s 는 다음과 같이 정의 된다. $s = k/n$ n 은 전체 데이터 개수를 가리키며, k 는 n 개중 선택할 데이터의 개수이다. k 값이 높아 질수록 s 값이 높아진다. s 값이 높을 수록 데이터를 많이 사용 하며 국부적인 평활화가 진행 되어 좀더 부드러운 평활화가 진행된다. 그에 반에 적은 s 값을 사용 할 수록 국소적으로 평활화가 진행되어 좀 더 구불한 평활화가 진행된다. 본 연구에 사용한 s 값은 0.7이다. 두 번째 x_0 에서 가까운 k 개의 데이터에 가중치 $Kx_0 = K(x_i, x_0)$ 를 할당하여 x_0 로부터 가장 가까운 점은 높은 가중치를 가지게 하고 가장 먼 점은 낮은 가중치를 가지게 한다. k 개를 제외한 모든 자료는 0의 가중치를 갖는다. 세 번째 Error를 최소화하는 $\hat{\beta}_0$ 과 $\hat{\beta}_1$ 을 구한다. 네 번째 x_0 에서의 적합 값을 $\hat{f}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 에 의해 계산한다. < 그림 12 > 은 $x_1(0.3 \sim 1.0$ 미세먼지) 데이터를 Simple Moving Average로 y (유량) 데이터는 Local Regression을 사용하여 노이즈를 제거한 데이터를 시각화한 결과이다.

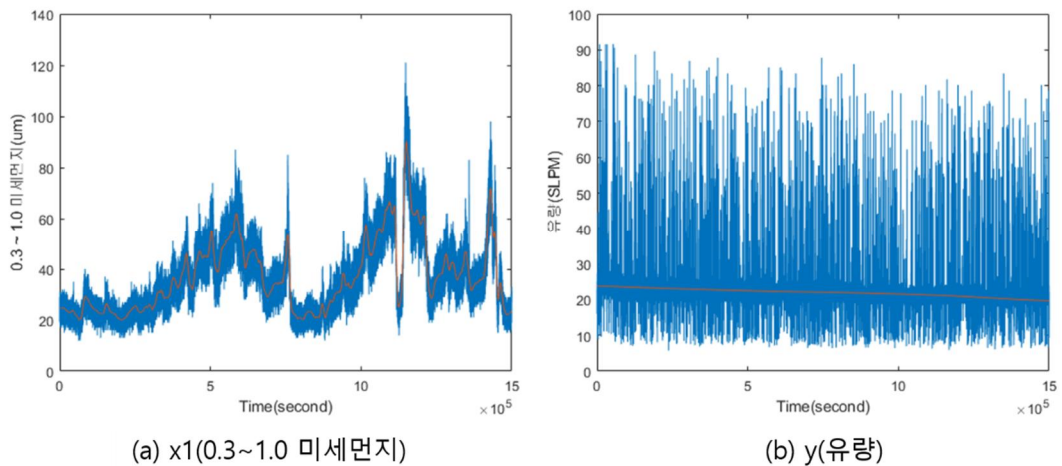


그림 12. 노이즈 제거 후 시각화

주황색 실선은 데이터에서 노이즈를 제거하고 실제 값으로 추출한 데이터이다. $x_1(0.3 \sim 1.0$ 미세먼지)의 경우 노이즈 제거를 통해 초단위로 변하는 값의 변화를 완화하였으며, y (유량) 데이터의 경우 사용 중 잘못 측정된 성능의 이상 변화를 제거하여 필터의 성능이 점차 감소하는 것을 확인 할 수 있다. 4.3.2 에 적용된 노이즈 제거를 통해 데이터의 특징이 담겨 있는 의미 있는 데이터를 추출 하여 분석 결과에 정확도를 높이도록 한다.

4.3.3 K-means 알고리즘을 이용한 환경 데이터 군집화

연속 값을 가지는 운용 환경 데이터(x_1, \dots, x_s)를 Clustering 기법을 사용하여 유사한 환경 간 그룹화를 진행한다. 이는 연속 값을 가지는 데이터 간 유사한 특성을 가지는 환경을 그룹화하여 환경 그룹의 특징을 분석하고 그룹이 가지는 특징에 따라 필터에 어떠한 영향을 가하는지 정량적으로 파악하기 위함이다.

본 연구에서 사용한 Clustering 기법은 군집 수를 지정하여 개별 데이터를 가장 가까운 군집으로 할당하는 K-means 알고리즘을 사용하였다. 알고리즘이 동작하는 원리는 네 단계로 구성된다. 첫번째 단계는 임의의 K개의 군집 개수를 정하여 랜덤 하게 k개의 데이터를 선택한다. 두번째 단계는 각각의 객체들을 가장 가까운 그룹의 데이터로 할당한다. 세번째 단계는 같은 그룹에 할당된 객체의 평균 값을 계산한다. 네번째 단계는 계산한 평균 값을 새로운 군집의 중심으로 할당한다. 마지막 다섯 번째 단계는 세 번째 네 번째 단계를 클러스터의 중심 값이 더 이상 변하지 않을 때까지 반복한다. 첫 번째 단계에서 최적의 군집 개수를 결정하기 위해본 논문에서는 DB(Davies-Bouldin)수치를 사용한다[44]. 군집수에 따른 DB수치는 더 작은 값을 가지는 경우가 더 우수한 성능으로 군집화가 수행됨을 의미하며 아래의 식으로 계산된다.

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

n 은 클러스터 개수이며 c 는 클러스터의 중앙값이며, σ 는 각 객체가 속해있는 클러스터의 중앙 값인 c 까지의 평균 거리이며, $d(c_i, c_j)$ 는 각 군집의 중앙값 간의 거리를 나타낸다. 분자에 있는 값은 각 객체가 속한 클러스터의 중앙값과의 거리이므로 클러스터내의 유사성이 높을 시 객체간 거리가 가까움으로 낮은 값을 가지게 된다. 분모의 경우 각 군집의 중앙값 간의 거리를 나타낸다. 이는 각 군집간의 거리가 멀수록 높은 값이 도출되며, 이는 클러스터간 유사성이 낮아 유사성이 깊은 객체간 군집화가 이루어 졌음을 알 수 있다. 즉, 객체가 속한 중앙값과 가까우며, 각 군집의 중앙값의 거리가 멀 수록 낮은 DB수치가 도출된다. < 표 3 >은 환경 데이터(x_1, \dots, x_5)를 변수 간 개별적으로 한 개의 군집부터 서른 개까지의 군집수로 군집화를 수행하여 DB 수치를 계산 한 것 중 가장 낮은 값을 도출하는 군집 개수 결과이다. 군집은 변수간 개별적으로 수행된다.

표 3. 최적의 군집 개수

Variable	Number of optimal k	Davies-Bouldin index
x_1 (0.3 ~ 1.0 미세먼지)	6	0.4359
x_2 (1.0 ~ 2.5 미세먼지)	6	0.4622
x_3 (2.5 ~ 10 미세먼지)	6	0.4272
x_4 (온도)	6	0.4683
x_5 (습도)	8	0.4449

<표 3>에 의하면 x_1 (0.3 ~ 1.0 미세먼지)의 최적의 군집은 여섯 개의 그룹이며 x_5 (습

도)는 여덟 개의 그룹으로 군집화를 수행 할 경우 가장 낮은 DB수치를 가지는 최적의 군집 개수를 나타낸다. <표 4>는 x_2 를 최적의 군집 개수로 군집화한 결과이다.

표 4. $x_2(1.0 \sim 2.5$ 미세먼지) 군집 결과

그룹	하한 값	중심 값	상한 값	할당 개수
1	21.5537	28.313	36.2187	263355
2	125.0602	147.9247	170.7471	471407
3	50.7801	57.4354	66.1708	274237
4	36.2189	44.1247	50.78	267799
5	88.551	102.1952	125.5083	27473
6	66.171	74.9065	88.5505	195729

< 표 4 > 에 따르면 그룹 1은 21.5537에서 36.2187의 사이에 존재하는 데이터가 할당 된다. 그룹 1의 중심 값은 28.313이며 471407개의 데이터가 그룹1에 할당되어 있으며, 그룹 6의 중심 값은 74.9065이며 263355개의 데이터가 그룹6에 할당 되었음을 확인 할 수 있다. < 그림 13 >은 $x_2(1.0 \sim 2.5$ 미세먼지) 군집화 결과를 시각화 한 것이다.

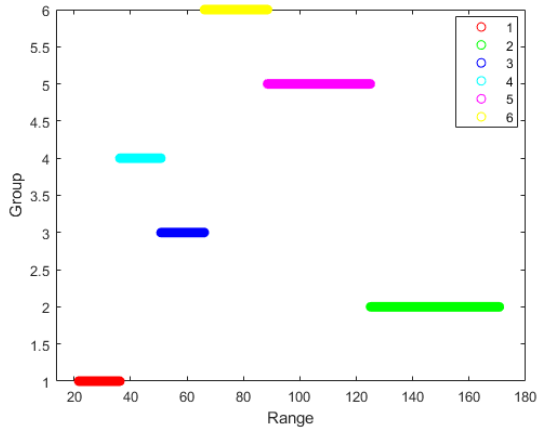


그림 13. 군집 시각화 x_2 (1.0 ~ 2.5 미세먼지)

< 그림 13 >에 x 축은 각 군집의 범위를 나타내며, y 축은 그룹을 나타낸다. 예를 들어 빨간색 선의 경우 y 축이 1인 곳에 존재한다. 이는 빨간색 선의 데이터는 1번 그룹에 할당된 데이터임을 의미하며, x 축은 그룹 1의 범위인 21의 하한 값에서 36의 상한 값 내의 데이터가 존재함을 의미한다. 4.3.4절에서는 연속 값을 가지는 데이터 간 유사한 특성을 가지는 환경을 그룹화하여 환경 그룹의 특징을 분석하였다. 이는 추후 분석 과정에서 환경 그룹이 필터 성능에 어떠한 영향력을 미치는지 정량적으로 파악할 수 있다.

4.3.4 예지 모델 구현을 위한 데이터 정의

본 절은 전처리된 데이터를 통해 필터가 운용 중인 환경인 x (input) 값과 운용 환경에 따른 필터 성능의 감소 값인 y (output) 값을 계산한다. 계산 방법은 네 단계로 구성되어 있다. 첫 번째 단계는 필터 운용 시작 시 가장 처음 측정되는 초기 유량 센서의 값을 기록한다. ($firts\ y = 23$) 두 번째 단계는 동일한 운용 환경이 연속으로 관측된 횟수를 기록한다. 예를 들어 운용 환경이 $(x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1)$ 인

환경이 40,000번이 반복 되었다면, ($firts\ y = 23, x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1$, 관측치 : 40000) 을 기록한다. 세 번째 단계는 반복 된 운용 환경에서 새로운 환경으로 변할 때의 유량 값을 기록한다. ($firts\ y = 23, x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1$, 관측치 : 40000, $second\ y = 22.7$) 기록된 데이터를 통해 운용 시작 시 23의 유량 값이 측정되었고 기록된 환경에서 40,000번의 필터가 노출되었으며, 그에 따라 제품의 운용 초기 유량에 비해 0.3의 유량 값이 감소된 것을 확인할 수 있다. 이를 백분율로 나타내면 ($\frac{22.7}{23} * 100\% = 98.69\%$)로 $(100\% - 98.69\%) = 1.31\%$ 의 성능이 감소되었음을 의미한다. 네 번째 필터가 환경에 한번 노출 될 경우 받는 성능 감소 값을 계산한다. 이는 $(x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1)$ 인 운용 환경에서 필터가 40,000번 노출되었을 경우 1.31%의 성능이 감소됨을 세 번째 단계를 통해서 알 수 있다. 따라서 $(\frac{1.31}{40000}) = 3.27E - 05$ 값은 운용 환경이 $(x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1)$ 인 경우 필터가 한번 노출 될 경우 받는 성능 감소 값이 되는 것이다. < 표 5 >는 수집된 환경에 필터가 한번 노출 될 경우 받는 성능 감소 값을 계산하였다.

표 5. 하중 (Input) 및 손상(damage) 데이터

$x_1(0.3 \sim 1.0$ 미세먼 지)	$x_2(1.0 \sim 2.5$ 미세먼 지)	$x_3(2.5 \sim 10$ 미세먼 지)	x_4 (온도)	x_5 (습도)	Y (성능 감소 값)
1	8	4	7	6	8.79e-06
1	8	4	8	6	1.57e-05
1	8	4	8	10	1.52e-05
⋮	⋮	⋮	⋮	⋮	⋮
6	9	6	7	2	6.04e-06
6	9	6	8	10	1.39e-05
6	9	6	10	4	9.65e-06

운용 환경에 따른 필터 성능 감소 값을 저장한 데이터 베이스의 첫 번째 열부터 세 번째 열까지는 운용 환경의 미세먼지 크기에 따른 먼지의 양을 k-means 알고리즘을 통해 그룹화한 그룹이 기록된다. 네 번째 열부터 다섯 번째 열까지는 k-means를 통해 그룹화된 온도와 습도의 그룹이 저장된다. 여섯 번째 열은 운용 환경 에서 필터가 한번 노출될 경우 필터 성능 감소 값을 기록한다. 필터에 가해지는 하중 (Input) 값은 $x_1 \sim x_5$ 의 운용 데이터이며 손상(damage) 데이터는 y 값이 되는 것이다.

4.3.5 학습 모델 정의

본 절은 수집되지 않은 필터의 운용 환경에 대응하는 성능 감소를 예측하기 위해 사용할 모델의 특성을 파악 한다. 본 연구에서 수집 가능한 환경의 경우의 수는 4.3.3 절에서 최적의 군집 개수를 모두 곱하여 계산한 결과 $(6 * 6 * 6 * 6 * 8) = 10,368$ 개의 환경이 존재하는 것을 알 수 있다. 그러나 본 연구에서 수집된 환경의 개수는

411개이며, 이는 수집되지 않은 9,957개의 새로운 환경이 발생할 시 새로운 환경에 대한 필터의 성능 감소 값을 파악 할 수 없는 문제점을 야기한다. 따라서, 운용 환경에 따른 필터 성능 감소 값의 상호 관계를 학습하여 운용 환경에 따라 실제 필터 성능 감소 값을 근사하게 예측 할 수 있는 예지 모델이 필요하다. 본 연구에 예측 값은 성능 감소 값으로 연속형 데이터로 표현된다. 이는 2.2.1에서 소개한 회귀 문제에 해당된다. 따라서 예측 값을 연속형 데이터로 출력하는 다양한 회귀 모델 사용이 필요하다.

본 연구에서 하중 대비 손상 데이터의 관계를 학습시키기 위해 사용된 방법은 서포트 벡터 회귀 모델(Support vector regression; SVR)과 딥뉴럴 네트워크(DNN) 그리고 선형 회귀 모델(Linear regression; LR)이다. 선형 회귀 모델은 종속변수를 독립변수와 오차 항의 선형결합으로 나타내는 방법이다. 한개의 독립 변수의 변화를 통해 종속 변수의 변화 량을 추정 할 경우는 단순 선형 회귀 분석 이며, 여러 개의 독립 변수의 변화를 통해 종속 변수의 변화 량을 추정 할 경우 다중 선형 회귀 분석이라고 표현한다. 독립 변수에 따른 종속 변수의 변화를 파악 하기 위해 실제 값과 예측 값의 차이인 잔차(residual)를 최소화 시키는 최소제곱법(least square estimation)을 사용하여 $y = ax + b$ 와 같이 일차 함수로 표현한다. 서포트 벡터 머신은 같은 출력 값을 가지는 데이터를 분류하기 다차원 공간 에서 데이터를 가장 잘 분류하는 최대마진 초평면(Maximum margin hyperplane)을 찾는 것이다. < 그림 14 >는 분류 직선을 찾아 분류하는 모델과 마진을 이용하여 분류하는 서포트 벡터 머신을 표현 하는 그림이다.

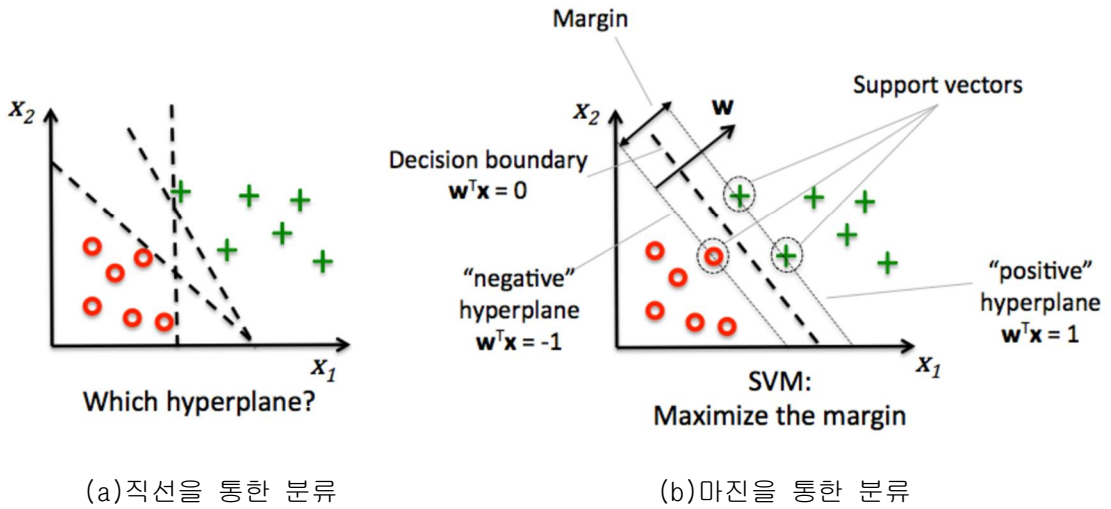


그림 14. Support vector machine

[그림 출처: <https://datascienceschool.net/view-notebook/6c6d450cb2ee49558856fd924b326e00/>]

두 가지 클래스를 분류하는 문제에서 (a)의 직선을 찾아 분류하는 경우 여러가지 방법이 존재한다. 이는 하나의 클래스의 쪽으로 선이 치중되어 있다면 새로운 입력 데이터에 대한 분류 정확도가 낮아 질 수 있다. 따라서 (b)의 클래스 간의 거리가 가장 큰 폭인 마진의 중간지점을 기준으로 분류할 경우 어느 한쪽으로 치중 되어있지 않은 분류 기준이 된다. 본 연구에서는 임의의 실수를 예측 할 수 있도록 분류 모델의 목적 함수를 변경하여 회귀 모델로 사용한다[43]. 딥뉴럴 네트워크는 2.1 에서 설명하였다.

4.3.5 절에서는 수집되지 않은 필터의 운용 환경에 대응하는 성능 감소를 예측하기 위해 사용할 세 가지 모델의 특성을 파악 하였다.

4.3.6 정규화 및 표준화

예지 모델을 통해 학습을 하기 위해서는 하중 (Input) $x_1 \sim x_5$ 데이터는 데이터의 정

규화(Normalization) 과정이 필요하다. 이는 다양한 센서에서 수집된 데이터의 단위 차이에 따른 편차가 분석에 주는 영향을 줄이기 위함이며, 데이터 범위를 일치 시키기 위해 수치형 변수를 변환하는 방법이다. 정규화는 학습 모델의 학습 단계의 속도를 높이며, 많은 연산 과정으로 인한 정확도 손실 문제를 예방하기 위해 사용한다 [54]. 본 연구에서는 가장 많이 사용되는 최대값과 최소값을 사용한 정규화를 사용한다. 본 방법은 데이터의 상한과 하한을 파악할 수 있을 때 사용하며, 정규화를 통해 두 개 이상의 서로 다른 단위의 데이터를 같은 기준으로 볼 수 있게 된다. 정규화 수행 이후 모든 센서 데이터는 종류별로 0에서 1 사이의 값을 가지게 되며 정규화에 사용된 식은 아래와 같다.

$$x_{(norm)i} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

$x_{(norm)i}$ 의 계산 방법은 표본 데이터에서 최대값과 최소값을 확인 하여 각 개별 데이터에서 최솟값을 빼준 데이터를 분자로, 최댓값에서 최솟값을 빼준 값을 분모로 하여 나누어 준다. < 표 6 >은 $x_1 \sim x_5$ 데이터가 정규화된 데이터 표이다. 정규화는 변수간 개별적으로 수행된다.

표 6. ($x_1 \dots x_5$) 정규화

x_1 (0.3 ~ 1.0 미세먼지)	x_2 (1.0 ~ 2.5 미세먼지)	x_3 (2.5 ~ 10 미세먼지)	x_4 (온도)	x_5 (습도)
0	0.1111	1	1	0.1111
0	0.7778	0.3333	0	0.4444
0	0.7778	0.3333	0.2222	1
⋮	⋮	⋮	⋮	⋮
0.8	0.4444	0.6667	0	0.3333
0.8	0.4444	0.6667	0	0.4444
0.8	0.4444	0.6667	0	0.5556

운용 환경 데이터($x_1 \dots x_5$)의 경우 그룹의 상한과 하한이 존재하여 최대값과 최소값을 사용하여 정규화를 수행한 반면 y (성능 감소 값)의 경우는 수집되지 않은 환경에서 필터를 운용 시 성능을 감소 하는 새로운 최대값과 최소값이 존재할 수 있다. 이는 현재까지 수집된 데이터를 통해 y (성능 감소 값)의 최대값과 최소값을 알 수 없다는 것이다. 따라서 y (성능 감소 값)의 경우 표준화(Standardization)를 사용한다. 표준화는 평균을 기준으로 데이터 값이 어느 정도 떨어져 있는지를 나타낼 때 사용하며, 사용된 식은 아래와 같다.

$$x_{(stand)i} = \frac{x_i - \mu}{\sigma}$$

$x_{(stand)i}$ 의 계산 방법은 표본 데이터에의 각 요소에서 평균을 뺀 다음 표준편차로 나누어준다. < 표 6 >은 운용 환경 $x_1 \sim x_5$ 데이터가 정규화를 y (성능 감소 값)이 표준화된 데이터 표이다.

표 7. ($x_1 \dots x_5$) 정규화 및 (y) 표준화

x_1 (0.3 ~ 1.0 미세먼지)	x_2 (1.0 ~ 2.5 미세먼지)	x_3 (2.5 ~ 10 미세먼지)	x_4 (온도)	x_5 (습도)	Y (성능 감소 값)
0	0.1111	1	1	0.1111	-0.8891
0	0.7778	0.3333	0	0.4444	0.0758
0	0.7778	0.3333	0.2222	1	-0.4474
⋮	⋮	⋮	⋮	⋮	⋮
0.8	0.4444	0.6667	0	0.3333	-0.6919
0.8	0.4444	0.6667	0	0.4444	-1.4097
0.8	0.4444	0.6667	0	0.5556	0.1345

정규화와 표준화를 수행한 데이터는 컴퓨터의 연산의 성능을 향상 시키기 위해 사용하며, 정규화 및 표준화된 데이터를 인간이 직관적으로 판단하기에는 무리가 있다. 따라서 이후 분석 단계에서 해석의 편리성을 주기 위해 표준화와 정규화를 수행 하기 전 데이터 값을 기록한다. 표준화의 경우 $x_i = (x_{(stdn)i} + \mu) * \sigma$ 로 정규화의 경우 $x_i = x_{(norm)i} * (\max(x) - \min(x)) * \min(x)$ 를 통해 표준화와 정규화를 수행 하기 전 데이터 값을 알 수 있다.

4.4 예측 모델 평가

본 절은 수집되지 않은 필터의 운용 환경에 대응하는 성능 감소 값을 예측하기 위해 여러 가지 모델을 사용하여 평가한다. 본 연구에서 수집 가능한 운용 환경의 경우의 수는 4.3.1 절에서 최적의 군집 개수를 모두 곱하여 계산한 결과 $(6 * 6 * 6 * 6 * 8) = 10,368$ 개의 환경이 존재하는 것을 알 수 있다. 그러나 본 연구에서 수집된 운용 환경의 개수는 411개이며, 이는 수집되지 않은 9,957개의 새로운 환경이 발생할 시 새로운 환경에 대한 필터의 성능 감소 값을 파악 할 수 없는 문제점을 야기한다. 따라

서 본 절에서는 수집된 411개의 운용 환경에 대응하는 필터 성능 감소 값의 상호 관계를 학습하여 수집되지 않은 9,957개의 운용 환경에 대응하는 필터 성능 감소 값의 예측을 목표로 한다. 본 연구에서 사용된 데이터의 상호 관계를 가장 잘 학습할 수 있는 우수한 성능의 모델을 선택하기 위해 본 연구에 사용된 세 가지 DNN, LR, SVR 모델에 대하여 Mean Absolute Percentage Error (MAPE)와 상관관계분석을 통해 비교한다. MAPE는 통계에서 예측 값의 정확도를 오차의 백분율로 측정하는 방법으로 사용된 식은 아래와 같다.

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

(A_t) 는 운용 환경에 따른 실제 성능 감소 값이고, (F_t) 는 운용 환경에 따른 실제 성능 감소 값을 모델이 예측한 값이며, n 개의 실제 값에서 n 개의 예측 값을 빼준 후 이를 실제 값으로 나누어준다. 계산된 모든 값을 더하여 n 으로 나누어 준다. 그 후 백분율로 표현하기 위해 100을 곱해준다. 값이 0에 근접할수록 모형 적합이 잘 되었다는 것을 의미한다. MAPE는 백분율로 표현되기 때문에 다른 정확도 척도 통계량 보다 쉽게 이해할 수 있는 장점이 있다. < 표 7 >는 본 연구에 사용된 세 가지 모델의 MAPE를 계산한 표이다.

표 8. MAPE

Model	MAPE
DNN	14.38
SVR	46.42
LR	45.66

MAPE 분석 결과 본 연구에 사용된 데이터를 가장 잘 설명하는 모델은 14.38의 MAPE 값을 가지는 DNN 모델이다. 다음은 운용 환경에 따른 실제 필터 성능 감소 값과 각 모델로 예측한 감소 값의 상호 관련성을 알아보기 위하여 상관관계 분석을 수행한다. 상관관계 분석은 두 변수가 서로 관련성이 있는지 추측하기 위해 사용하는 기법이며 상관관계의 정도를 나타내는 수치를 상관계수(r)라 한다. 상관관계는 한 변수의 값이 증가할 때 다른 변수의 값이 같이 증가하는 경우 양의 상관관계라고 하고, 한 변수의 값이 증가할 때 다른 변수의 값이 감소하는 경우 음의 상관관계라고 한다. 두 변수 사이 어떠한 관계도 없는 경우에는 무 상관이라고 한다. 상관계수는 항상 $-1 \leq r \leq 1$ 을 만족시키며, 양의 상관관계는 $r > 0$, 음의 상관관계가 있을 때는 $r < 0$ 이다. 무 상관이 경우는 $r = 0$ 근처에 가까운 값이 되며, 0.6 이상이거나 -0.6이하인 경우 두 변수 간 강한 상관관계를 가지고 있다고 말하고 있다. 실제 필터 성능 감소 값과 모델로 예측한 필터 성능 감소 값의 상관관계를 계산한 결과는 < 표 8 >과 같다.

표9. 상관 계수

Model	r
DNN	0.8374
SVR	0.1958
LR	0.1985

< 표 8 >의 상관계수 결과에 따르면, 실제 성능 감소 값과 모델을 통해 예측된 값의 가장 높은 상관계수는 0.83이다. 추가적으로 상관관계가 통계적으로 높은 수치인 0.7 보다 큰 상관계수를 가지고 있는지 검증 하기 위해 T-test를 사용하여 가설검정을 수행한다. T-test의 유의 수준은 엄격한 검정을 위해 0.01로 한다. T-test의 귀무가설은 $H_0: p > 0.7$ 로 환경에 따라 감소되는 실제 필터 성능 감소와 모델을 통해 예측된 필터 성능 감소의 상호 관계가 0.7 이상의 높은 상관계수를 가지고 있다는 것이고, 대립가설은 $H_1: p < 0.7$ 로 실제 값과 예측 값의 상호 관계가 0.7 이상의 높은 상관계수를 가지고 있지 않다는 것이다.

표 10. T-test

$\alpha = 0.01$	DNN	SVR	LR
Test statistic	3.25	-3.08	-3.06
Test results	H_0 기각 못함	H_0 기각	H_0 기각

< 표 9 >에 따르면 DNN모델의 검정 결과는 유의수준 0.01에서 운용 환경에 따른 실제 성능 감소 값과 운용 환경에 따라 DNN을 통해 예측한 성능 감소 값이 0.7이상의 높은 상관계수를 가지는 것을 알 수 있다. SVR과 LR모델의 경우에는 운용 환경에 따른 실제 필터 성능 감소 값과 각 모델을 통해 예측한 필터 성능 감소 값이 0.7이상의 높은 상관계수를 가지지 않는다는 결과가 나타났다. 즉, 운용 환경에 따라서 DNN모델을

통해 예측한 필터 성능 감소 값 이 실제 운용 환경에 따른 필터 성능 감소 값과 유사성이 있다는 것이다. 검정 결과를 통해 운용 환경에 따른 실제 필터 성능 감소 값과 운용 환경에 따라서 DNN모형을 통해 예측된 필터 성능 감소 값, 두 값 사이에 상호연관성이 존재함을 확인 할 수 있었다. < 그림 14 >는 환경에 따른 실제 운용 환경에 따른 필터 성능 감소 값과 각 모형을 통해 운용 환경에 따라 예측한 필터 성능 감소 값을 그린 산점도 그림이다.

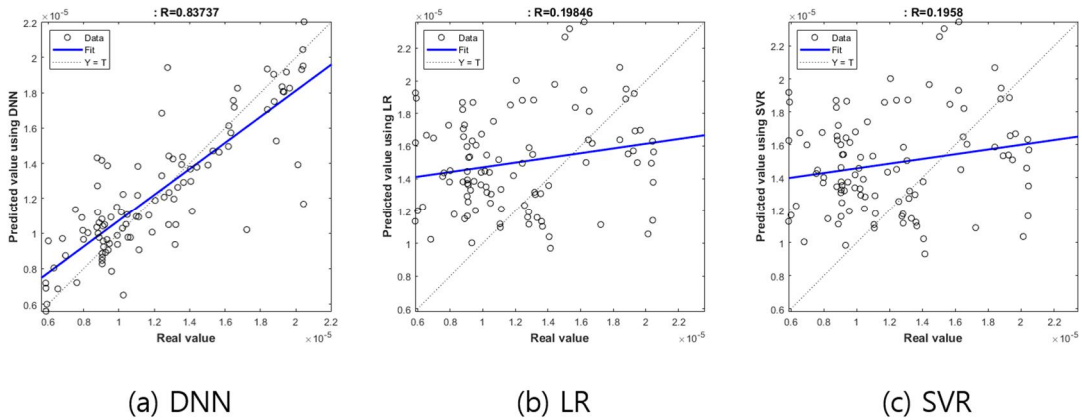


그림 15. Scatter of real value and predicted value

시각화를 통해서도 실제 운용 환경에 따른 필터 성능 감소 값 과 DNN모형을 통해 예측한 운용 환경에 따른 필터 성능 감소 값의 상호연관성이 가장 높음을 알 수 있다. 따라서 본 연구에 데이터를 가장 잘 설명 하는 모델은 14%의 가장 낮은 MAPE 값과 0.83의 상관계수를 가지는 DNN모형을 알 수 있다. < 표 10 >은 DNN모형을 사용하여 수집되지 않은 9,957개의 환경에 따른 성능 감소 값을 예측한 결과이다.

표 11. 미수집 운용 데이터에 대한 필터 성능 감소 값 예측

Variable	Combination1	Combination2	Combination9956	Combination9957
x_1 (0.3 ~ 1.0 미세먼지)	1	1	1	5
x_2 (1.0 ~ 2.5 미세먼지)	1	1	3	5
x_3 (2.5 ~ 10 미세먼지)	1	1	1	6
x_4 (온도)	2	1	3	4
x_5 (습도)	7	6	2	3
Y(성능 감소 값)	3.3825e-05	3.3625e-05		5.7037e-06	5.6964e-06

< 표 10 >의 결과 운용 환경이 ($x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 2, x_5 = 7$)인 경우에 필터가 한번 노출되면 3.3825e-05의 성능이 감소한다는 것을 알 수 있다. 수집되지 않은 9,957개의 환경에 대응하는 필터 성능 감소를 예측하였다. 따라서 수집된 411개의 환경과 DNN모델을 통해 예측된 9,957개를 더한 값인 10,368개의 환경에 대응하는 필터 성능 감소 데이터를 확보하였다. 10,368개의 환경 중 제품 운용 중 제품에 성능을 가장 빠르게 감소 시키는 최악의 환경과 제품의 성능이 가장 오랫동안 유지되는 최상의 환경을 파악한다. 이는 제품 운용 환경에 따른 최악의 상황과 최상의 환경을 사전에 파악하여 설비의 운용능력을 보장하기 위함이다. < 표 11 >은 각 환경에서 필터가 운용 될 경우 성능이 0에 도달 할 때까지의 소모 시간을 적은 것부터 많은 순으로 나열한 것이다.

표 12. 필터 수명 범위

Variable	Combination1	Combination2	Combination10367	Combination10368
x_1 (0.3 ~ 1.0 미세먼지)	1	1	4	6
x_2 (1.0 ~ 2.5 미세먼지)	5	4	5	6
x_3 (2.5 ~ 10 미 세먼지)	6	6	6	6
x_4 (온도)	1	1	3	2
x_5 (습도)	1	6	5	2
<i>RUL(day)</i>	57	57		403	404

< 표 12 >의 결과 운용 환경이 ($x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1$)인 경우에 필터가 지속적으로 운용 될 경우 제품의 성능이 0에 도달할 때까지 시간이 57일 소요된다는 것이다. 이를 통해 최악의 경우인 57일 전 후로 필터의 유지보수를 시행 할 경우 최상의 설비 운용 능력을 보장 할 수 있다. 가장 오랜 시간 제품의 성능이 보장되는 경우는 운용 환경이 ($x_1 = 6, x_2 = 6, x_3 = 6, x_4 = 2, x_5 = 2$)인 경우에 운용 시 404일 동안 사용이 가능하다. 따라서 필터를 최대 404일 이내에는 교환 해줘야 한다. < 그림 15 >는 각 환경에서 필터가 운용 될 경우 성능이 0에 도달 할 때까지의 소모 시간을 시각화 하였다.

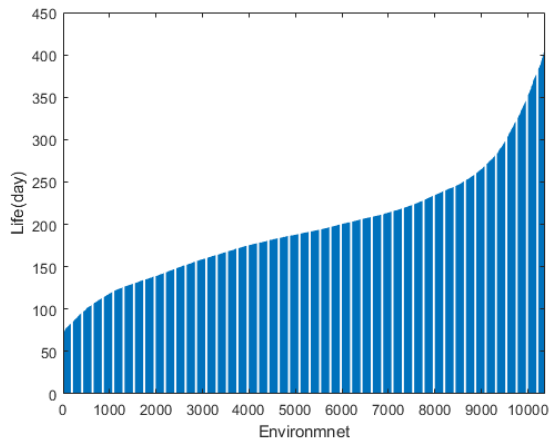


그림 16. 제품 수명 범위

MAPE 계산을 통해 본 연구에 사용된 데이터를 가장 잘 설명 할 수 있는 모델은 DNN 이다. 모델에 구현된 DNN 파라미터는 다음과 같다. 심층 신경망을 학습 시키기 위한 네트워크로는 입력 층으로 데이터가 입력 되며, 은닉층을 거쳐 출력 층으로 출력 값을 보내는 Feed-forward 네트워크를 사용하였다. 레이어는 입력층과 은닉층 그리고 출력층 총 3개로 구성되어있다. 입력층의 neuron 개수는 본 연구에 수집된 입력 데이터 (x_1, \dots, x_5) 개수인 5개 이며, 은닉층의 layer는 세 개로 구성 되어있다. 첫 번째 은닉층은 relu activation function을 가지는 열 개의 neurons 이 존재 하며 두 번째 은닉층은 sigmoid activation function을 가지며, 일곱 개의 neurons이 존재한다. 마지막 세 번째 은닉층은 tanh activation function을 가지는 여덟 개의 neurons이 존재한다. 출력층은 임의의 실수를 출력하는 하나의 neuron이 존재한다. 트레이닝 알고리즘은 Back-propagation 알고리즘을 사용하였다[8]. 트레이닝 최적화 알고리즘은 Levenberg에 의해 고안 되고 Marquardt가 보완한 Levenberge-Marquardt 알고리즘을 사용하였다[47,48]. 학습을 파라미터인 Learning rate는 값이 높은 경우 최소값을 찾지 못하여 무한대로 발산하게 되며, 매우 낮은 경우 학습이 매우 느리게 진행된다.

하지만 Learning rate에 대한 최적의 값을 계산 하는 방법에 대한 이론적인 방법은 아직 존재하지 않으며, 그에 따라 많은 연구들이 진행되고 있다[49]. 본 연구에 사용된 모델의 Learning rate는 0.1로 설정하였다. 모델의 Cost Function은 가장 많이 사용되는 Mean squared error(MSE)로 설정하였다. 이는 예측 값과 실제 실제 값 과의 차이를 제공해서 모두 더한 값들의 평균으로 정의된다. 심층 신경망을 학습 하는 것 중 가장 큰 문제중 하나는 오버피팅 이다. 이는 학습 데이터 셋에 맞는 최적의 모델로 학습하여 추후 예측해야하는 일반 적인 데이터 셋에 대한 예측 정확도가 낮음을 의미한다. 따라서 이를 해결하기 위해 모델의 설명도를 유지하며 복잡도를 줄이기 위해 함수의 곡률을 줄여주는 역할을 하는 Regularization 값을 사용한다. 본 연구에 사용된 값은 0.2로 설정하였다. 마지막 파라미터인 데이터 셋의 분류는 모델의 학습에 사용되는 Training set 이 60%의 비율을 차지하며, 학습된 모델의 성능을 검증하며 파라미터를 수정하는Validation set은 15%의 비율을 차지한다. 마지막 모델의 정확도를 측정하기 위한 Test set은 25%로 나누어 학습을 진행하였다. < 표 12 >는 모델 파라미터에 대한 값을 기록하였다.

표 13. DNN모델 파라미터

Parameter	Description
Network	Feed-forward
Number of Layer	Three
Input layer	Five neurons
Output layer	One neurons
Hidden layer	Ten neurons - relu Seven neurons - sigmoid Eight neurons - tanh
Training Algorithm	Backpropagation
Optimization Algorithm	Levenberg-Marquardt
Learning Rate	0.1
Cost function	Mse
Regularization	0.2
Percentage of Data utilized in Training	60%
Percentage of Data utilized in Validation	15%
Percentage of Data utilized in Testing	25%

이번 단계에서 DNN모형을 통해 수집되지 않은 9,957개의 환경에 대응하는 필터 성능 감소 값을 예측하였다. 추가적으로 필터 운용 중 성능 감소가 가장 빠르게 일어나는 환경($x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1$)과 가장 오랫동안 설비의 운용 능력을 보장하는 환경 ($x_1 = 6, x_2 = 6, x_3 = 6, x_4 = 2, x_5 = 2$)을 파악하였다. 다음 단계에서는 Bootstrap 기법을 사용하여 제품의 평균 잔여유효수명을 예측한다.

4.5 Bootstrap을 이용한 평균 잔여유효수명 예측

4.4절을 통해 가장 오랫동안 설비의 운용 능력을 보장하는 환경과 가장 짧은 설비의 운용 능력을 보장하는 환경을 파악하였다. 하지만 시간이 지남에 따라 변화하는 환경에서 같은 환경이 지속적으로 운용되는 경우는 매우 드물다. 따라서 제품의 평균 잔여유효수명을 추정할 필요가 있다. 따라서 필터 성능이 소모되는 평균 잔여유효수명을 사전에 파악하여 최적의 필터 교체 시점을 지정하여 유지 보수를 수행하는 방법이 필요하다. 이를 위해 본 연구에서는 Bootstrap 기법을 사용한다. Bootstrap 기법은 추정 하고자 하는 데이터의 확률분포 및 신뢰구간을 이론적으로 구하기 힘든 경우에 사용 한다. 사용 방법은 주어진 데이터 $X = (x_1, x_2, x_3, \dots, x_N)$ 중 n 개의 데이터를 i 번 리샘플링한다. 그 후 평균, 분산 및 신뢰구간등 구하고자 하는 파라미터 값을 계산한다. 예를 들어 구하고자 하는 파라미터가 평균이고 주어진 데이터가 $X = (x_1, x_2, x_3, \dots, x_4)$ 집합이며, 리샘플링 할 데이터의 n 을 3개로 리샘플링 반복 횟수 i 인 i 를 10으로 가정하였을 경우 $X_1 = (x_2, x_4, x_3) \sim X_{10} = (x_4, x_1, x_2)$ 의 데이터를 수집한다. 그후
$$\text{mean} = \frac{(\text{mean}(x_1) + \text{mean}(x_2) + \dots + \text{mean}(x_{10}))}{10}$$
 과 같이 평균을 계산한다. 본 연구에서는 제품의 평균 수명을 추정 하기 위해 사용한다. Bootstrap에 사용되는 데이터 집합은 운용 환경에 따른 필터 성능 감소 값으로 설정한다. 데이터 집합에서 리샘플링 되는

데이터는 운용 환경이 나올 확률을 기반으로 추출된다. 리샘플링할 데이터의 개수인 $n = 1$ 로 설정한다. 리샘플링 반복 횟수 i 는 제품의 성능이 0이 되는 시점까지 이다. Bootstrap 리샘플링시 환경이 나올 확률은 네 단계에 걸쳐서 계산한다. 첫 번째 단계는 운용 환경 데이터(x_1, \dots, x_5)의 각 변수 별 환경 그룹 분포를 파악한다. 두 번째 단계는 각 변수 별로 군집화 된 환경 그룹의 상대 도수를 각 그룹 환경이 나올 확률로 정의한다. < 그림 16 >은 x_2 (1.0~2.0 미세먼지)의 분포를 나타낸다.

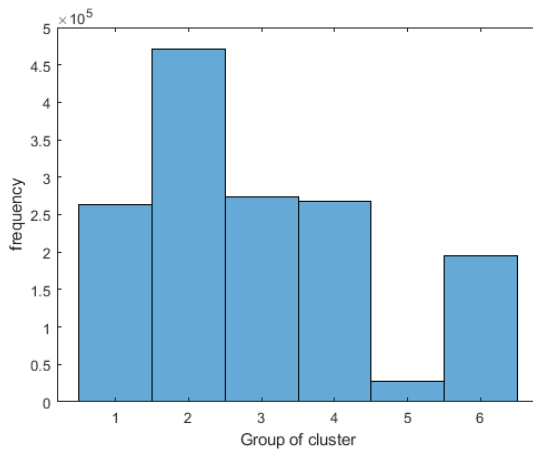


그림 17. x_2 (1.0~2.0 미세먼지)의 분포

x_2 (1.0~2.0 미세먼지)는 6개의 환경 그룹이 존재한다. 150만개의 데이터 중 첫 번째 그룹은 263,355개의 데이터가 존재하며, 따라서 $\left(\frac{263.355}{1500000} = 0.1756\right)$ 의 상대 도수를 가지며, x_2 (1.0~2.0 미세먼지)의 첫 번째 그룹이 나올 확률은 17.56%이다. 이와 같은 방법으로 환경 변수의 각 그룹이 가질 확률을 계산해준다. 세 번째 단계는 운용 환경 데이터(x_1, \dots, x_5)의 모든 조합인 10,368개의 운용 환경이 나올 확률을 계산한다. 예를 들어 두 번째 단계를 통하여 각 변수 별 그룹이 가질 확률을 계산하였다면

$(p(x_1 = 1) = 10\% , p(x_2 = 1) = 17\%, p(x_3 = 6) = 20\%, p(x_4 = 4) = 7\%, p(x_5 = 5) = 3\%)$
 같은 조합이 본 연구의 운용 환경 데이터의 조합인 10,368개가 존재 한다. 이 데이터
 셋의 운용 환경 안에 존재하는 그룹 안에 있는 확률 값을 곱해준다.
 $(p(x_1 = 1) = 10\% , p(x_2 = 1) = 17\%, p(x_3 = 6) = 20\%, p(x_4 = 4) = 7\%, p(x_5 = 5) = 3\%)$
 의 경우 $(0.1 * 0.17 * 0.2 * 0.07 * 0.03) = 0.00000714$ 의 값이 나오며, 총 10,368개의 값
 을 계산 할 수 있다. 그리고 계산된 10,368개의 값을 더해준다. 그후 10,368개의 값
 을 더한 값을 분모로, 각 조합의 그룹 안에 있는 확률 값을 곱해준 값을 분자로 놓고
 계산 해준다. 이와 같은 과정을 수행하면 각 운용 환경이 나올 확률이 계산된다. 모
 든 환경이 나올 확률을 더하면 1이된다. 그 후 각 환경이 나올 확률을 기반으로
 Bootstrap을 사용하여 100%의 필터 성능이 소모되는 시간의 평균 수명을 추정한다. <
 표 13 > 은 본 연구에서 수집한 데이터와 4.4절에서 DNN모델을 사용하여 예측한 데이
 터를 사용하여 각 운용 환경 에서 2초간 운용 시 감소되는 성능 값과 운용 환경이 나
 올 확률을 계산한 표 이다.

표 14. Bootstrap을 위한 데이터 집합

Variable	Combination1	Combination2	Combination10367	Combination10368
x_1 (0.3 ~ 1.0 미세먼지)	1	1	4	6
x_2 (1.0 ~ 2.5 미세먼지)	5	4	5	2
x_3 (2.5 ~ 10 미 세먼지)	6	6	6	6
x_4 (온도)	1	1	3	2
x_5 (습도)	1	1	5	2
y (성능 감소 값)	4.02e-05	10		5.73e-06	5.72e-06
P	2.70e-05	2.63e-05	1.14e-05	1.36e-05

< 표 13 >의 두번째 열은 ($x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 2, x_5 = 7$) 필터가 한번 노출되면 4.02e-05의 성능 감소 값을 가지는 것을 의미한다. 일곱 번째 행은 각 운용 환경이 나올 확률로 이를 기반으로 샘플링 과정에서 추출된다. < 표 13 > 기반으로 하여 Bootstrap기법을 사용하며, 100%의 필터 성능이 소모되는 평균 수명을 추정한다. Bootstrap은 두가지 가정을 기반으로 수행된다. 첫 번째 샘플링은 각 환경이 관측될 확률을 기반으로 수행된다. 두 번째 각 환경은 복원 추출을 기반으로 수행된다. 두가지 가정을 기반으로 Bootstrap은 다음과 같이 수행 된다. 첫 번째 단계는 각 환경에서 한번 노출시 감소되는 필터 성능 값을 집합 X 로 정의한다. 두 번째 단계는 주어진 $X = (x_1, x_2, x_3, \dots, x_{10368})$ 중 리샘플링 할 데이터의 n 을 1개로 리샘플링 반복 횟수 i 를 제품의 성능이 100%가 감소 될 때까지로 한다. 반복 횟수 i 를 측정하여 제품의 잔여 유효수명으로 계산한다. 본 연구의 센서는 2초에 한번 기록된다. 따라서 한번의 i 는 2초로 계산된다. 예를 들어 반복 횟수 i 가 천만 번이 기록되었을 때 제품의 성능이 0

이되었다면, 제품 수명은 이천 만초 즉, 231일이었다는 뜻이다. 세 번째는 두 번째 과정을 10번 반복하여 필터의 평균 수명을 추정한다. < 표 >는 10번의 Bootstrap을 통해 얻어진 수명 데이터이다.

표 15. Bootstrap을 통해 계산된 필터 평균 수명

Sample	Life(time)	Life(day)
1	15056398	174
2	15052778	174
⋮	⋮	⋮
9	15056768	174
10	15054572	174
mean	15055307	174

운용 환경 분포를 고려하여 제품의 평균 수명을 계산한 결과 필터의 평균 수명이 174일로 추정 된다. 이번 단계에서 수행하는 Bootstrap을 통해 운용 환경의 분포를 고려하여 필터의 평균 수명을 예측 하였다. 4.3 - 4.5를 통해서 필터 운용 환경에 따라서 필터의 성능 감소 값이 다를 수 있었으며, 사용 수명이 달라 짐을 확인할 수 있다. 가장 오래 필터의 성능을 보장 할 수 있는 기간은 404일 이며, 가장 짧은 성능을 보장 할 수 있는 기간은 57일 이다. 또한 운용 환경의 분포를 고려하여 Bootstrap을 사용하여 제품의 잔여유효수명을 추정한 결과 평균 수명은 174일로 추정 된다. 센서의 성능이 발달하여 더욱 정밀한 측정이 이루어지며, 더욱 많은 데이터가 수집된다면 더욱 정확한 수명이 추정 가능 할 것이며, 최적의 유지보수를 통해 높은 신뢰도를 제고할 것이다.

제 5 장 결론

본 논문에서 제안된 방법론은 운용 중인 공기청정기의 필터에서 수집된 센서 데이터를 사용하여 운용 환경에 따른 필터의 성능 감소를 분석하는 것이다. 제안된 방법론을 통해 다양한 운용 환경에 따른 성능 감소를 계산하고 제품의 성능이 가장 짧게 운용 되는 환경과 가장 길게 운용 되는 환경을 찾을 수 있다. 이를 위해 제안된 방법론은 Deep neural network와 Clustering 같은 기계학습 기법과 Local regression 등의 다양한 통계적 기법을 사용하고 있다. 본 연구에서 분석 대상이 된 필터에 제안된 방법론을 적용하여 특정 환경 상태($x_1 = 1, x_2 = 5, x_3 = 6, x_4 = 1, x_5 = 1$)에서 운용 할 경우 57일의 수명을 가지며 가장 빠른 필터의 성능 소모가 발생하는 것으로 확인되었다. 그에 반해 가장 긴 필터 수명 주기를 가지는 운용 환경은($x_1 = 6, x_2 = 6, x_3 = 6, x_4 = 2, x_5 = 2$)에서 운용 할 경우 제품의 필터 성능이 404일간 보장 됨을 알 수 있다. 또한 필터의 운용 환경 분포를 고려하여 Bootstrap을 통해 필터의 평균 수명을 추정한 결과 174일의 평균 수명을 가짐을 확인 할 수 있다. 본 연구를 통해 예측의 정확도를 높여 필터의 최적 교체 시점을 체계적으로 찾을 수 있다.

본 연구에서 제안된 방법론은 운용 환경에 따른 제품의 성능 감소 값을 사용하여 제품의 수명을 추정했다는 점에서 다양한 운용 환경에 적용 될 수 있다.

하지만 데이터의 수집 기간이 35일로 다소 짧아 더 많은 데이터를 수집하여, 예지 모델의 성능을 한층 성장 시켜 더 정확한 결과를 도출해야 하는 부분이 필요하다.

향후 연구에서는 제품의 노화 정도에 따라 제품의 성능 감소 값이 다르게 받을 수 있는바 운용 환경과 함께 필터의 노화 정도에 따라서 필터의 성능 값도 계산하여 연구의 결과를 더욱 증진 시킬 것 이다.

참고문헌

1. LG CNS, <http://blog.lgcns.com/1250>
2. 김인석, 장양, 정원., 2015, 최근 항공장비의 고장유형과 예방보전 정책의 변화, *한국신뢰성학회 학술대회논문집*, pp. 339-344.
3. Samuel, AL., 1959, Some studies In machine learning using the game of checkers, *IBM Journal of Research and Development*, Vol. 3(3), pp.201-229.
4. Turing, A. M., 2009, Computing machinery and intelligence, *Parsing the Turing Test*. Springer, Dordrecht, pp.23-65.
5. McCulloch, W. S. and Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Vol. 5(4), pp.115-133.
6. Rosenblatt, F., 1958, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological review*, Vol. 65(6), 386-408.
7. Minsky, M. and Papert, S. A., 2017. Perceptrons: An introduction to computational geometry, MIT press.
8. Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1985, Learning internal representations by error propagation (No. ICS-8506), *California Univ San Diego La Jolla Inst for Cognitive Science*.

9. Baly, R. and Hajj, H., 2012, Wafer classification using support vector machines, *IEEE Transactions on semiconductor manufacturing*, Vol. 25(3), pp.373-383.

10. Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., and An, W., 2018, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, *Automation in Construction*, Vol. 85, pp.1-9.

11. Butler, S., & Ringwood, J., 2010, Particle filters for remaining useful life estimation of abatement equipment used in semiconductor manufacturing, In *Control and Fault-Tolerant Systems (SysTol)*, pp. 436-441.

12. Kankar, P. K., Sharma, S. C., & Harsha, S. P., 2011, Fault diagnosis of ball bearings using machine learning methods, *Expert Systems with applications*, Vol. 38(3), pp.1876-1886.

13. Heng, A., Tan, A. C., Mathew, J., Montgomery, N., Banjevic, D., and Jardine, A. K., 2009, Intelligent condition-based prediction of machinery reliability, *Mechanical Systems and Signal Processing*, Vol. 23(5), pp.1600-1614.

14. 삼성 sds icto

15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014. Generative adversarial nets, *In Advances in neural information processing systems*, pp.2672-2680.

16. Ahmad, R., and Kamaruddin, S., 2012, An overview of time-based and condition-based maintenance in industrial application, *Computers & Industrial Engineering*, Vol. 63(1), pp.135-149.

17. Luo, J., Pattipati, K. R., Qiao, L., and Chigusa, S., 2008, Model-based prognostic techniques applied to a suspension system. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 38(5), pp.1156-1168.

18. Wang, H., 2002, A survey of maintenance policies of deteriorating systems. *European journal of operational research*, Vol. 139(3), pp.469-489.

19. Dasgupta, A., and Pecht, M., 1991, Material failure mechanisms and damage models, *IEEE Transactions on Reliability*, Vol. 40(5), pp.531-536.

20. Tinga, T., 2010, Application of physical failure models to enable usage and load based maintenance, *Reliability Engineering & System Safety*, Vol. 95(10), pp.1061-1075.

21. Qiu, J., Seth, B. B., Liang, S. Y., and Zhang, C., 2002, Damage mechanics approach for bearing lifetime prognostics. *Mechanical systems and signal processing*, Vol. 16(5), pp.817-829.

22. Gu, J., Barker, D., and Pecht, M., 2007, Prognostics implementation of electronics under vibration loading. *Microelectronics reliability*, Vol. 47(12), pp.1849-1856.

23. Wilkins, E. W. C., 1956, Cumulative damage in fatigue. In *Colloquium on Fatigue*, Springer, Berlin, Heidelberg, pp. 321-332.

24. Oppenheimer, C. H., and Loparo, K. A., 2002, Physical based diagnosis and prognosis of cracked rotor shafts. In *Component and Systems Diagnostics, Prognostics, and Health Management II*, Vol. 4733, pp. 122-133.

25. Kacprzynski, G. J., Sarlashkar, A., Roemer, M. J., Hess, A., & Hardman, B., 2004, Predicting remaining life by fusing the physics of failure modeling with diagnostics, *JOm*, Vol. 56(3), pp.29-35.
26. Chen, L., Liu, Y., and Xie, L., 2007, Power-exponent function model for low-cycle fatigue life prediction and its applications-Part II: Life prediction of turbine blades under creep-fatigue interaction, *International journal of fatigue*, Vol. 29(1), pp.10-19.
27. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J., 2018, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, *Mechanical Systems and Signal Processing*, Vol. 104, pp.799-834.
28. Cubillo, A., Perinpanayagam, S., and Esperon-Miguez, M., 2016, A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery, *Advances in Mechanical Engineering*, Vol. 8(8), pp.1-21.
29. Jardine, A. K., Lin, D., and Banjevic, D., 2006, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical systems and signal processing*, Vol. 20(7), pp.1483-1510.
30. Wang, Z., and Wang, P., 2012, Reliability-based product design with time-dependent performance deterioration, In *Prognostics and Health Management (PHM), 2012 IEEE Conference on*, pp. 1-12.
31. Vogl, G. W., Weiss, B. A., and Helu, M., 2016, A review of diagnostic and prognostic capabilities and best practices for manufacturing, *Journal of Intelligent Manufacturing*, pp.1-17.
32. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., and Siegel, D., 2014, Prognostics and health management design for rotary machinery systems-Reviews, methodology and applications, *Mechanical systems and signal processing*,

Vol. 42(1-2), pp.314-334.

33. Medjaher, K., Tobon-Mejia, D. A., and Zerhouni, N., 2012, Remaining useful life estimation of critical components with application to bearings, *IEEE Transactions on Reliability*, Vol. 61(2), pp.292-302.

34. Heng, A., Tan, A. C., Mathew, J., Montgomery, N., Banjevic, D., and Jardine, A. K., 2009, Intelligent condition-based prediction of machinery reliability, *Mechanical Systems and Signal Processing*, Vol. 23(5), pp.1600-1614.

35. Si, X. S., Wang, W., Hu, C. H., and Zhou, D. H., 2011, Remaining useful life estimation-a review on the statistical data driven approaches. *European journal of operational research*, Vol. 213(1), pp.1-14.

36. Sikorska, J. Z., Hodkiewicz, M., and Ma, L., 2011, Prognostic modelling options for remaining useful life estimation by industry, *Mechanical Systems and Signal Processing*, Vol. 25(5), pp.1803-1836.

37. Ho, S. L., Xie, M., and Goh, T. N., 2002, A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction, *Computers & Industrial Engineering*, Vol. 42(2-4), pp.371-375.

38. Gong, S. F., He, Y., and Jin, W. L., 2007, Fatigue life reliability analysis of single point mooring offshore jacket platform structures, *JOURNAL-ZHEJIANG UNIVERSITY ENGINEERING SCIENCE*, Vol. 41(6), pp.995.

39. Hu, C., Youn, B. D., and Chung, J., 2012, A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation, *Applied Energy*, Vol. 92, pp.694-704.

40. Tang, D., Yu, J., Chen, X., and Makis., V., 2015, An optimal condition-based maintenance policy for a degrading system subject to the competing risks of soft and hard failure, *Computers & Industrial Engineering*, Vol. 83, pp.100-110.
41. Walsh, D. C., and Stenhouse, J. I. T., 1997, The effect of particle size, charge, and composition on the loading characteristics of an electrically active fibrous filter material. *Journal of aerosol science*, Vol. 28(2), pp.307-321.
42. 장영민, 정재희, 이관형, 차영식, 정춘기, and 이덕출., 1996, 대전방지처리된 물질의 정전기 대전특성에 관한 연구. *Journal of KIIS*, Vol. 10(1).
43. Gunn, S. R., 1998, Support vector machines for classification and regression. *ISIS technical report*, Vol. 14(1), pp. 5-16.
44. Davies, D. L., and D. W. Bouldin., 1997, A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No. 2, pp. 224-227.
45. Botsaris, P. N., & Tsanakas, J. A., 2008, State-of-the-art in methods applied to tool condition monitoring (TCM) in unmanned machining operations: a review. In *Proceedings of the Int. Conference of COMADEM, Prague*, pp. 73-87
46. Haldar, A., & Mahadevan, S., 2000, Probability, reliability, and statistical methods in engineering design, Vol. 1
47. Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, Vol. 2(2), pp. 164-168.
48. Marquardt, D. W., 1963, An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, Vol. 11(2), pp. 431-441.
49. Smith, L. N., 2017, Cyclical-learning rates for training neural networks. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pp.

464-472

50. James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013, An introduction to statistical learning, Vol. 112, New York: springer.
51. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., and Beghi, A., 2015, Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, Vol .11(3), pp. 812-820.
52. Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., and Hampapur, A., 2014, Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, Vol. 45, pp. 1726.
53. Liao, S. H., 2005, Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert systems with applications*, Vol. 28(1), pp. 93-103.
54. Liu, Z., 2011, A method of SVM with normalization in intrusion detection. *Procedia Environmental Sciences*, 11, 256-262.

감사의 글

학부 과정을 시작으로 대학원 석사 과정 동안 많은 지식을 주시고, 공학도가 가져야 할 마음 가짐과 성실함을 항상 몸소 보여주시며, 부족한 저를 교수님의 첫 번째 대학원생으로 받아주시고, 연구에 매진할 수 있도록 항상 많은 도움을 주신 저의 지도 교수님 이신 신종호 교수님께 진심으로 감사 드립니다. 또한 저의 논문 심사를 맡아 주시고, 많은 조언과 도움을 주신 백정우 교수님, 김성준 교수님께 감사드립니다. 그리고 대학 생활 동안 많은 조언을 해주신 김규태 교수님, 강장하 교수님, 박형준 교수님께도 깊이 감사드립니다.

SF.Lab 첫 구성원으로 시작하여 대학원 석사 과정 동안 함께 있어준 연구실 후배들에게도 항상 미안하고 고맙다는 말을 전합니다.

또한 사랑하는 어머니와 먼 곳에서 지켜보고 계실 아버지 그리고 어릴 적부터 항상 챙겨주신 이모 정말 사랑하고 감사합니다. 또한 가장 오랜 벗 상훈, 성환 에게도 고맙다는 말을 전합니다.

마지막으로 감사의 글을 통해 언급하지 못했지만 항상 저의 주변에서 함께 해주시는 모든 분들께 다시 한번 진심으로 감사 드립니다.