기계학습을 이용한 이상진단 기술에 관련된 이슈

이승철*, 민형철, 정해동
(울산과학기술대학교 인간 및 시스템 공학부)

1. 머리말

최근 각광받는 분야인 빅데이터와 밀접한 관련이 있는 기계학습(machine learning)에 기반한 이상진단 기술과 적용방법에 대해 실제 사례를 통하여 간략히 소개한다. 또한 이를 적용하는데 있어서 현재 어떠한 이슈들이 논쟁이 되는지, 그리고 이를 해결하기 위하여 어떠한 방법이 필요한지 향후 연구 방향을 제시하고자 한다.

ICT(information & communication technology), 빅데이터(big data), IoT(internet of things) 등 방대한 양의 데이터에서 사용자에게 유용한 정보를 추출하는 방법론이 주목 받고 있다. 데이터 처리 방법론 중 가장 널리 알려진 기계학습은 인공지능의 한 분야로, 컴퓨터가 학습할 수 있도록 알고리듬과 기술을 개발하는 분야를 말한다. 기계에 여러 센서를 부착하여 데이터를 취득하는 이상진단 기술도 특성상 빅데이터의 한 분야로 볼 수 있다. 이에 착안하여 생긴 방법론이 바로 기계학습을 이용한 이상진단 기술이다. 기존의 기계의 시스템적 해석을 통하여 이상진단을 실시하는 방법과는 달리 기계학습을 이용하여 이상진단을 실행하는 방법으로 센서로 취득된 데이터를 통계적인 처리와 확률론적 모델에 기반 하여 현재 기기의 상태가 어떤지, 향후 기기의 성능저하상태가 어떻게 진행되는지를 파악하는 방법이다.

2. 기계학습에 기반한 이상진단

기계학습에 기반한 이상진단이 기존의 동적해석을 통한 이상진단과 가장 큰 차이점은 시스템의 동적해석이 아닌 데이터로 설비 상태를 판단한다는 점이다. 이상진단 기술에 통계를 접합함으로써 복잡한 기계시스템을 해석할 필요 없이 데이터에 기반하여 기계학습 알고리듬을 구축한 뒤, 모니터링 시스템이 스스로 설비의 건강 상태를 파악할 수 있도록 할 수 있다. 이상진단에 기계학습을 적용하는 절차는 그림 1과 같이 4단계로 나눌 수 있다.

(1) 생산라인에서 중요자산을 선정 한 뒤, 선정된 기기에 센서를 달고 데이터를 취득하는 단계이다. 이 과정에서는 중요자산을 어떻게 선정할 것인가, 센서를 어느 위치에, 몇 개를 달아야 하는가가 중요하게 다뤄진다.

(2) 취득된 데이터를 DSP를 통하여 특성신호로 추출하는 단계이다. 특성신호를 선정하는 과정은 기계학습에 기반한 이상진단에서 가장 중요하다고 해도 과언이 아닐 정도로 매우 중요한 절차이다. 시스템의 동적 특성을 반영하는 특성신호를 선정하면 효율적으로 데이터의 정보 손실 없이 크기를 줄일 수 있으며 다음에 이어질 기계학습 알고리듬을 통한 기계상태의 판단에도 효과적으로 사용될 수 있다.

* E-mail : seunglee@unist.ac.kr

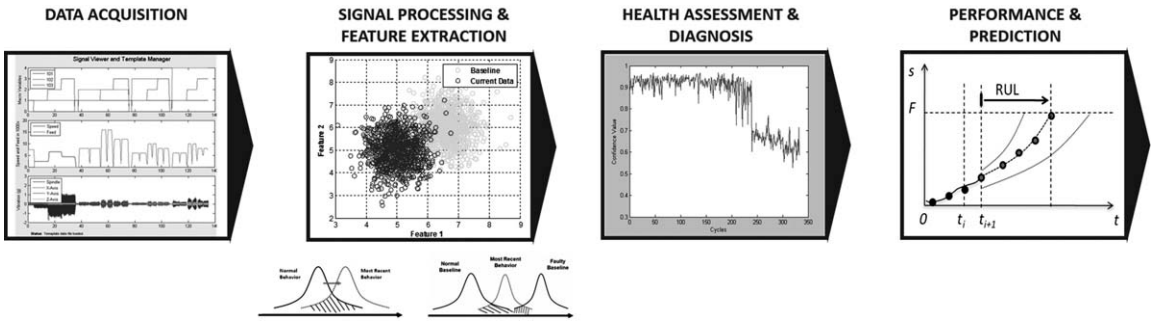


그림 1 기계학습을 통한 이상진단 방법론

표 1 기계학습 관점에서의 해석

질문	기계학습 관점
어떠한 데이터를 받아야 하는가?	Input (센서의 종류)
설비의 어느 위치에 부착해야 하는가?	Input (센서의 위치)
센서를 몇 개 부착해야 하는가?	Multivariate data와 input dimension의 문제
신호를 그대로 사용할 것인가?	Feature extraction and selection

(3) 알고리듬을 통하여 기기의 상태를 진단하는 과정이다. 데이터에 라벨이 있을 경우 지도학습(supervised learning), 그렇지 않을 경우 자율학습(unsupervised learning)으로 분류되며 각각 적용될 수 있는 알고리듬의 종류가 다르다. 크게 분류(classification)의 방법을 사용하여 기기의 고장모드를 판별하는 방법, 회귀분석(regression)을 사용하여 기계의 열화상태를 추적 및 예측하는 방법이 있다.

(4) 확률론적 방법을 이용한 기기 상태의 예측 및 잔존유효수명(remaining useful life) 예측이다. 현재 기기상태를 기반으로 향후 기기의 상태가 어떤 식으로 변해갈지 확률론적 기법을 통해 예측한다.

3. 이상진단과 관련된 이슈

인공지능 시스템에 정상 상태의 신호를 학습시킨 뒤 알고리듬에 따라 시스템이 스스로 기기의 이상상태를 진단하도록 하는 것이 기계학습에 기반한 이상진단의 기본 구조이다. 언뜻 보기에 복잡한 동적 해석 필요 없이 기계학습을 이용한 기기의 이상진단은 모든 것을 가능하게 하는 만능열쇠처럼 여겨질 수 도 있다. 하지만 기계학습에 기반한 이상진단의 결과를 맹목적으로 신뢰하는 것은 문제가 있을 수 있다. 특성신호를 선정할 때 시스템의 동적 특성을 반영하는 특성신호(feature)를 선정할 지더라도 데이터 기반(data-driven)의 통계 방법상 그 한계가 명백히 존재하기 때문이다. 따라서 기계학습을 구현하더라도 그 결과를 받아들이는데 있어서 고려할 사항이 존재한다. 이와 관련된 이슈를 소개 한다.

3.1 센서의 종류, 센서의 위치, 센서의 개수, 특성신호

데이터 취득 시스템(DAQ: data acquisition system)에 센서의 종류, 위치, 개수 등은 기계학습에 기반한 이상진단 시스템을 구축하는데 있어서 가장 먼저 접하게 되는 문제이며, 동시에 진단 성능에 가장 많은 영향을 미치는 단계이다. 표 1과 같은 질문은 기계학습 관점에서 해석될 수 있다.

특성신호를 선정할 때는 현장 전문가의 의견을 반영하여 센서의 종류, 위치, 개수를 정하는 방법에 더하여, 그림 2와 같이 기계학습 방법을 사용

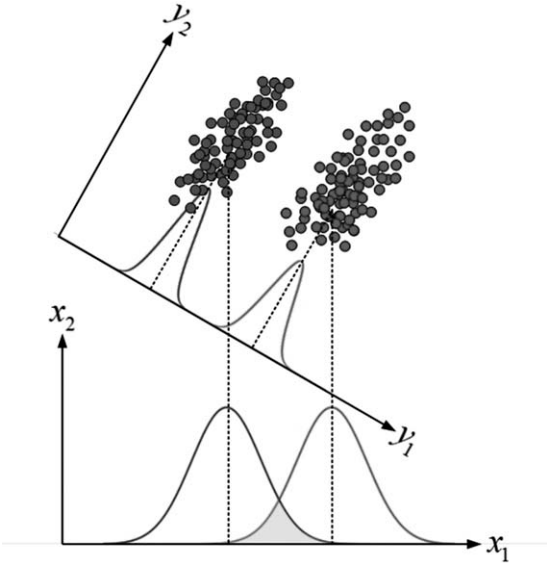


그림 2 특성신호 선택의 중요성

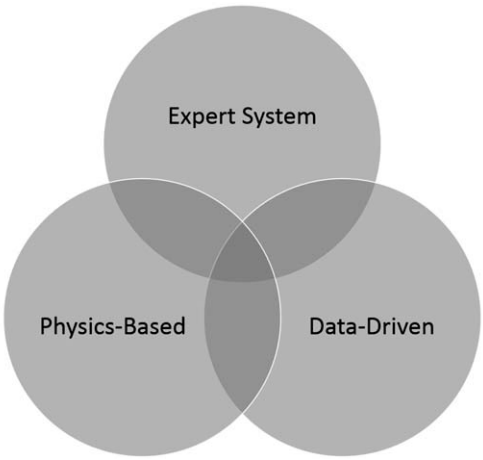


그림 3 진단의 3요소

하면 유용하다. 예를 들어, 같은 신호라도 x축으로 신호를 보면 서로 다른 클러스터가 겹쳐있어서 구별을 정확하게 할 수 없지만, y축으로 신호를 처리하면 서로 다른 클러스터로 명확하게 구별 할 수 있다. 이렇듯 기계학습은 진단에 가장 유리한 특성신호를 선정하는데 도움이 된다.

(1) 고장 신호를 알기 위해 기계를 일부러 고장내야 한다?

통계적 확신을 가지고 설비 상태를 판단하기 위해서는 정상상태의 신호뿐만 아니라 고장상태의 신호도 데이터베이스에 많이 축적이 되어 있어야 한다. 하지만 비정상 신호의 경우 생산라인의 고장이 발생할 경우가 거의 없고, 고장이 났었더라도 그 신호 데이터의 기록이 남아있지 않는 경우가 대부분이다. 이로 인해 고장신호의 축적을 위해 기기를 일부러 고장내야 하는 모순점이 발생한다. 유사한 시스템이나 비교적 작은 테스트베드를 통해 연구를 많이 하고 있지만, 데이터를 기반으로 하는 특성 때문에, 경향성을 보여줄 수는 있어도 시스템이 달라지면 정확한 진단을 할 수 없게 된다. 이러한 문제점들을 보완하기 위해 그림 3과 같이 설비 설계자 혹은 현장 전문가의 의견을 진단에 포함시키는 경우가 많으며, 다소 개발 시간이 걸릴 수는 있지만 모델링(physics-based model)을 통하여 비정상 신호의 패턴을 시뮬레이션을 통해 재현하는 연구도 이루어지고 있다.

(2) 상관관계와 인과관계를 파악

기계학습을 이상진단에 적용할 경우 그림 4(a)처럼 X_1, X_2, ..., X_n개의 특성신호(또는 인자, 프로세스 파라미터)와 결과 y의 상관관계를 파악하는 알고리듬을 적용하는 것이 대부분이다. 인자들과 결과의 연관성은 두 인자간에 선형성이 얼마나 강한지를 나타내는 상관관계(correlation)으로 표현된다. 상관관계는 통계적 공분산 행렬(covariance matrix)의 계산을 통해 구할 수 있다. 일반적으로는 인자들간의 구조를 무시하고 x와 y의 상관관계를 규명하는데 그친다. 하지만 실제의 시스템의 경우 인자들은 그림 4(b)와 같이 다양한 구조로 형성되어 있다. 시스템을 근본원인분석(root cause analysis)적 관점에서 보았을 때, 인자들 간의 상관관계뿐만 아니라 인과관계를 규명하는 것이 중요하다. 다만 인자들 간의 구조를 데이터에서 찾아내는 것은 복잡하고 현실적으로 어려운 문제이다. 그럼에도 불구하고 인자들 간의 구조와 인과관계를 파악 할 수 있으면 이는 진단에 상당히 도움이 된다. 그림 4(b)와 같은 구조에서 결과 y에 문제가 있으면,

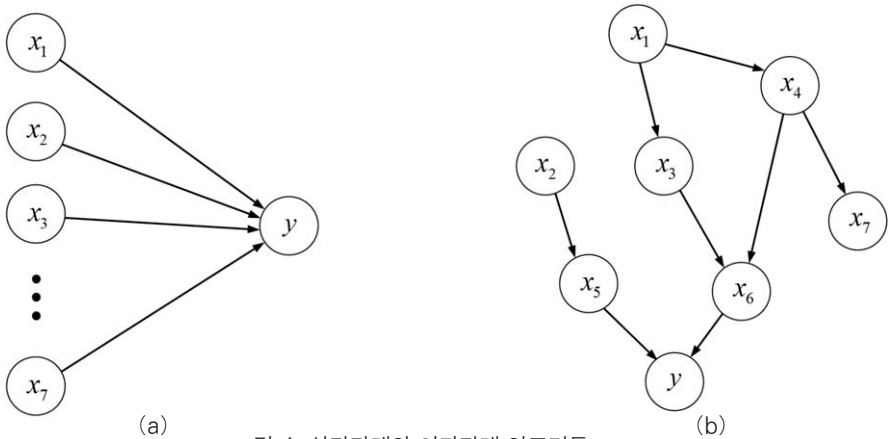
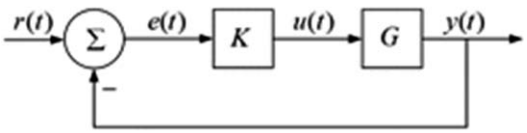


그림 4 상관관계와 인과관계 알고리듬



$$Y(s) = \frac{GK}{1+GK} R(s)$$

그림 5 피드백 제어 시스템

많은 인자 중에 X_5와 X_6을 가장 먼저 조사하여야 한다.

(3) 피드백 시스템은 좋은 것인가?

자동화 설비는 대부분 제어기술이 적용되어 있다. 음의 피드백 제어(negative feedback control)를 이용하면 그림 5의 수식에서 볼 수 있듯이 출력 신호가 기준 신호를 따라간다. 제어를 통해, 노이즈나 외란(disturbance)이 존재하더라도 원하는 시스템의 출력 신호를 유지할 수 있다. 하지만 기계학습을 통한 진단의 관점에서는 시스템 G가 변하여도 출력 신호 $y(t)$에서 그 변화가 미비하게 관찰되기 때문에 설비의 정확한 상태를 모니터링 하는데 제어는 오히려 방해가 된다. 피드백 제어가 적용되어 있는 설비에 진단 모듈을 새롭게 추가하려고 한다면, 반드시 이 점을 고려해야 한다.

(4) 유지 보수가 기계의 특성을 바꾼다

현장에서는 정기적으로 설비를 유지 보수한다. 정기적인 점검을 통해 기계의 조기 이상 검출이나 소모품 등을 교체하는 것이 목적인데, 이 유지 보수가 시스템의 특성을 바꿀 수도 있다. 사전에 학습된 데이터를 통한 이상진단 방법은 유지 보수로 인한 시스템의 동특성 변화를 반영하는 것이 아니므로 문제가 될 수 있다.

4. 기계학습을 통한 이상진단 기술의 실례

실제로 이상진단을 하는데 있어서 기계학습이 어떻게 적용되는지 그림 6의 회전체 테스트베드를 이용해 설명한다.

(1) 궤적의 특성신호를 이용한 고장검출

취득된 x, y신호를 통해 축중심(shaft)의 궤적을 생성한다. 필터를 통해 추출된 1X 성분의 궤적(orbit)은 타원형을 이룬다. 타원형 궤적은 forward response와 reverse response 벡터의 합과 위상차로 이루어지며, 이는 풀스펙트럼(full spectrum) 분석을 통해 구할 수 있다. 회전상태(정상, 비정상, 고장 유형, 다른 RPM)에 따라 궤적의 형태가 달라지기 때문에 그림 7처럼 타원의 장축, 단축의 길이, 위상을 특성신호로 하는 기계학습을 통해 회전체의 진단이 가능하다.

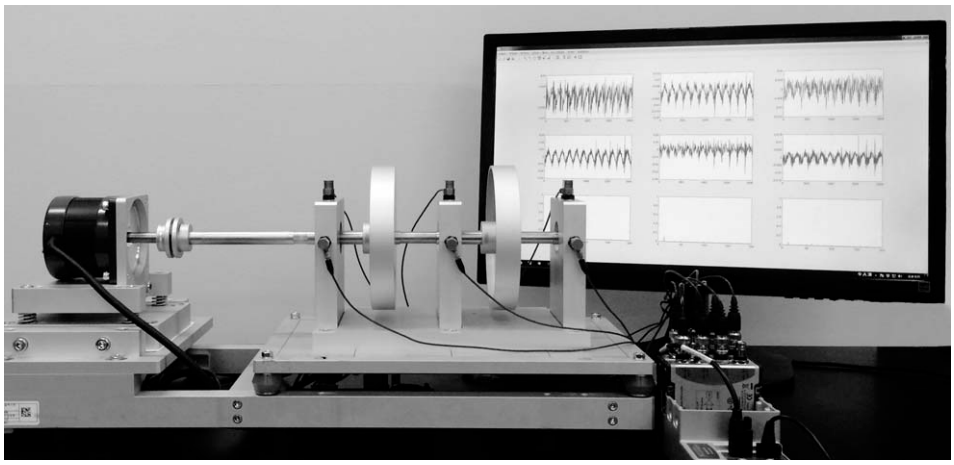


그림 6 회전체 시스템

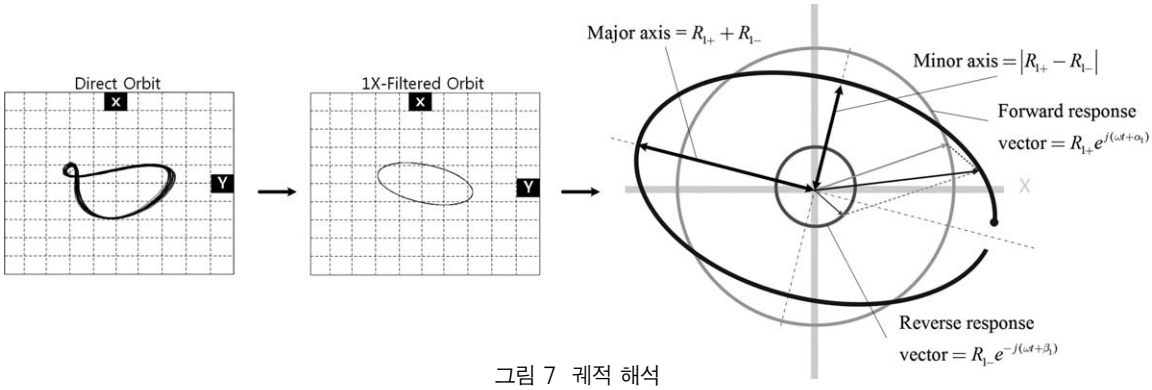


그림 7 궤적 해석

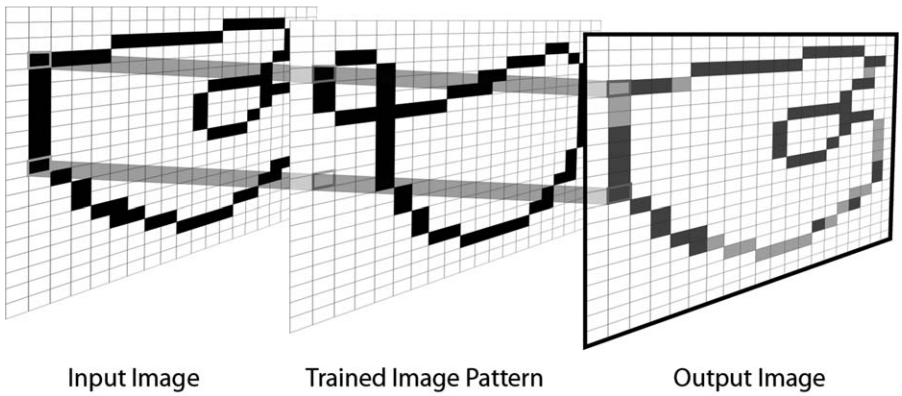


그림 8 궤적 이미지 패턴인식

(2) 이미지 패턴인식을 활용한 고장검출

궤적의 경우 특성신호를 추출하기보다 자체 형태가 패턴이 되어 설비 상태를 나타내므로, 현장에서는 궤적의 패턴을 보고 사람이 회전체의 상태를 자주 판단한다. 궤적의 패턴을 이미지화하여, 컴퓨터비전에서 얼굴 인식을 위해 사용되는 기계학습 방법을 그림 8과 같이 진단에 적용할 수 있다. 실시간으로 생성되는 궤적을 미리 학습된 이미지 패턴과 비교하여 설비의 회전 상태를 진단한다.

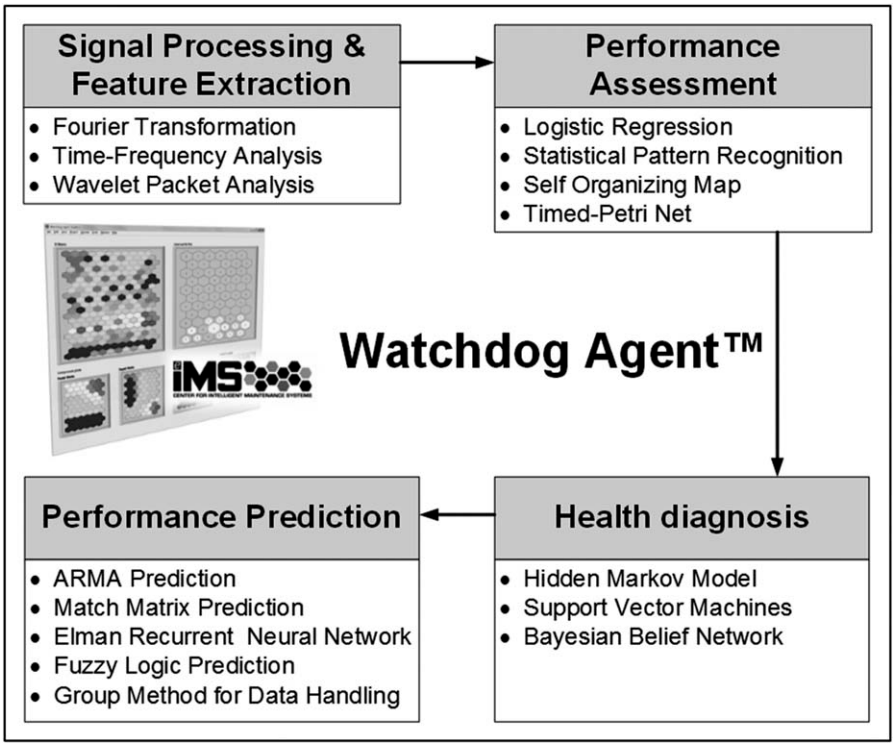


그림 9 Watchdog Agent™

5. Watchdog Agent™의 알고리듬 기술을 활용한 이상진단 시스템 구축

Watchdog Agent™은 저자가 2006년부터 2013년까지 몸담았던 미국 IMS(Intelligent Maintenance Systems) 센터와 National Instruments 사가 함께 LabVIEW기반으로 개발한 툴 킷으로서 이상진단에 필요한 주요 알고리듬을 한 곳에 모아둔 것이다. 기본적인 DSP와 feature extraction, performance assessment, health diagnosis와 performance prediction에 필요한 알고리듬이 포함되어있다.

6. 향후 전망

데이터 기반 기계학습의 한계점을 극복하기 위한 몇 가지의 연구방향을 제시하고자 한다.

첫 번째, '진화하는 DB'의 개념을 제안한다. 기계학습에 기반한 이상진단은 학습에 필요한 데이터베이스에 많이 의존하게 된다. 따라서 고장유형에 따른 신호의 확보가 힘들 경우에는 기계학습을 사용하기 힘들다. 따라서 먼저 자율학습의 anomaly detection을 통하여 정상 신호와 비정상 신호를 분리한다. Anomaly detection은 기기의 고장모드 분류에는 사용할 수 없지만 기기의 정상상태와 비정상상태를 구분할 수 있는 알고리듬이다. 설비의 운전시간이 늘어남에 따라 비정상상태도 발생할 것이며, 이를 학습하고 DB에 추가함으로써 기계학습의 한계점을 극복하고자 하는 것이 하나의 방향이다.

두 번째, 데이터베이스들 간의 결합이다. 실제 산업현장에서는 신호기반 DB 뿐만 아니라 규칙기반(rule-based) DB도 사용된다. 일반적인 규칙과 현장 전문가들에 의해 축적된 규칙을 DB화해서 이상진단에 이용한다. 이 두 방법론(또는 DB)을 결합해서 데이터를 분석한다면 보다 더 정확한 진단을 내릴 수 있을 것이다. KSNVE