Journal of Next-generation Convergence Information Services Technology Vol.7, No.2, December (2018), pp. 141-148
차세대컨버전스정보서비스기술논문지 *http://dx.doi.org/10.29056/jncist.2018.12.02*

AI 머신러닝을 이용한 자동차 고장진단 시스템 설계 및 개발

Design and Development of Automobile Fault Diagnosis System Using AI Machine Learning

배재환[1]

Jae-Hwan Bae[1]

요 약

본 논문에서는 기계학습을 이용한 자동차 고장진단 시스템 설계에 관한 것으로, 설치위치에서 주행 중인 자동차의 소음을 검출하는 다수의 음향센서를 포함하여 부품 이상에 따라 발생되는 소음데이터를 검출하여 전송하는 소음검출부와, 상기 소음검출부에서 검출된 소음데이터를 외부로 전송하며, 수신된 정보를 표시하는 사용자 단말과, 상기 소음데이터를 분석하여 고장 부품의 종류와 위치를 특정하고, 고장 부품의 종류에 따라 위험도를 결정하고, 소음데이터의 크기와 발생빈도를 이용하여 고장 정도를 판단하여 수리 지급도를 결정하여, 상기 고장 부품의 종류, 위치, 위험도, 수리 지급도 정보를 상기 사용자단말로 전송하는 진단 서버를 포함 한다. 이를 통해서 기계학습 관련 분야 산업 발전에 도움이 되었으면 한다.

핵심어 : 기계학습, 인공지능, 딥러닝, 고장진단시스템

Abstract

The present invention relates to a design of an automotive fault diagnosis system using machine learning, and more particularly, to a vehicle fault diagnosis system using a machine learning system including a noise detector for detecting and transmitting noise data generated by a component abnormality including a plurality of acoustic sensors, A user terminal for transmitting the noise data detected by the noise detecting unit to the outside and displaying the received information; a noise analyzing unit for analyzing the noise data to identify the type and position of the faulty component, determine the risk according to the type of the faulty component, And a diagnosis server for determining the degree of repair by determining the degree of repair using the size and frequency of noise data, and transmitting the type, position, risk, and repair degree information of the failed component to the user terminal. I hope this helps to develop industry related to machine learning.

Keyword : Machine learning, Artificial intelligence, Deep running, Fault diagnosis system

1 Department of Game Engineering, TongMyong University, Busan, Korea [Professor]
e-mail: bjhmail@tu.ac.kr

* This paper has been studied by the support of the academic research funding in the school of Tongmyung University in 2018 (Project number / 2018F003)

Received(October 24, 2018), Review Result(1st: November 18, 2018), Accepted(December 04, 2018), Published(December 31, 2018)

ISSN: 2384-101X JNCIST
ⓒ 2018 NCISS

1. 서론

최근 국내외적으로 구글 알파고 때문에 인공지능(AI)에 대한 관심이 매우 높다. 본 논문에서는 인공지능의 대표적인 방법이었던 전문가 시스템은 사람이 직접 많은 수의 규칙을 집어넣는 것을 전제로 하였다. 이 같은 접근 방법은 과학에 기 반한 학문들, 예를 들어 의학이나 생물 분야에서는 큰 역할을 할 수 있었다. 의사들의 진단을 도와주는 전문가 시스템에 기 반한 프로그램을 생각해 보면 인간이 지금까지 발견한 의학적인 규칙들을 데이터베이스화 하여 등록시켜주면 되는 것이었다[1]. 기계학습의 정의는 어떠한 태스크(T)에 대해 꾸준한 경험(E)을 통하여 그 T에 대한 성능(P)를 높이는 것, 이것이 기계학습이라고 할 수 있다. 정의에서 알 수 있듯이, 기계학습에서 가장 중요한 것은 E에 해당하는 데이터이다[1][2]. 이에 본 논문에서는 머신러닝을 이용한 자동차 고장진단 시스템 설계를 제안 하고자 한다. 논문의 주요 구성은 1장 서론, 2장 기계학습 분류, 3장 고장진단시스템 설계, 4장 결론으로 구성 된다.

2. 기계학습 분류

기계학습의 정의는 어떠한 태스크(T)에 대해 꾸준한 경험(E)을 통하여 그 T에 대한 성능(P)를 높이는 것, 이것이 기계학습이라고 할 수 있다. 정의에서 알 수 있듯이, 기계학습에서 가장 중요한 것은 E에 해당하는 데이터이다. 기계학습 문제들을 학습 종류에 따라 3가지로 나눌 수 있다. 특히, 레이블(label)의 유무에 따라 지도학습과 비지도 학습으로 나뉘는데, 여기서 레이블이란, 학습 데이터의 속성을 무엇을 분석할 지에 따라 정의되는 데이터를 뜻한다[3][4].

2.1 지도 학습(Supervised Learning)

사람이 교사로 써 각각의 입력(x)에 대해 레이블(y)을 달아놓은 데이터를 컴퓨터에 주면 컴퓨터가 그것을 학습하는 것이다. 사람이 직접 개입하므로 정확도가 높은 데이터를 사용할 수 있다는 장점이 있다. 대신에 사람이 직접 레이블을 달아야 하므로 인건비 문제가 있고, 따라서 구할 수 있는 데이터양도 적다는 문제가 있다.

① 분류(Classification): 레이블 y가 이산적(Discrete)인 경우 즉, y가 가질 수 있는 값이 [0,1,2 ..]와 같이 유한한 경우 분류, 혹은 인식 문제라고 부른다. 일상에서 가장 접하기 쉬우며, 연구가 많이 되어있고, 기업들이 가장 관심을 가지는 문제 중 하나다. 이런 문제들을 해결하기 위한 대표적인 기법들로는 로지스틱 회귀법[5], KNN, 서포트 벡터 머신 (SVM), 의사 결정 트리 등이 있다.아래와 같은 예시가 있다.

ⓒ 2018 NCISS

· 주차게이트에서 번호판 인식: 요새 주차장들은 티켓을 뽑지 않고, 차량 번호판을 찍어서 글자를 인식하는데 정확도를 높인다. 번호판른 정형화 되어있으므로 전통적인 컴퓨터 비전으로도 처리는 가능하나, 오염 등에 대해 정확도를 높이자면 기계학습을 하면 더 좋다. 이미지 픽셀 값들에 따라 숫자 글자를 분류한다.

· 페이스북이나 구글 포토의 얼굴 인식: 역시 컴퓨터 비전을 이용하되 기계학습을 결합. 페이스북에 사진을 올리면 친구 얼굴 위에 이름이 자동으로 달리고는 하는데, 이것 역시 기계학습을 이용한 것. x가 이미지 픽셀, y가 사람 이름인 경우.

· 음성 인식: 음성 wav 파일에 대해서 해당 wav 부분이 어떤 음절인지를 인식하는 것. 애플 시리, 구글 보이스 등에서 사용된다(질문에 대해서 답해주는 부분 말고, 인식 부분만). x가 음성 파형, y가 음절.

② 회귀(Regression): 레이블 y가 실수인 경우 회귀문제라고 부른다. 보통 엑셀에서 그래프 그릴 때 많이 접하는 바로 그것이다. 데이터들을 쭉 뿌려놓고 이것을 가장 잘 설명하는 직선 하나 혹은 이차함수 곡선 하나를 그리고 싶을 때 회귀기능을 사용한다. 잘 생각해보면 데이터는 입력(x)와 실수 레이블(y)의 짝으로 이루어져있고, 새로운 임의의 입력(x)에 대해 y를 맞추는 것이 바로 직선 혹은 곡선이므로 기계학습 문제가 맞다. 통계학의 회귀분석 기법 중 선형회귀 기법이 이에 해당하는 대표적인 예이다.

2.2 비지도 학습(Unsupervised Learning)

사람 없이 컴퓨터가 스스로 레이블 되어 있지 않은 데이터에 대해 학습하는 것. 즉 y없이 x만 이용해서 학습하는 것이다. 정답이 없는 문제를 푸는 것이므로 학습이 맞게 됐는지 확인할 길은 없지만, 인터넷에 있는 거의 모든 데이터가 레이블이 없는 형태로 있으므로 앞으로 기계학습이 나아갈 방향으로 설정되어 있기도 하다. 통계학의 군집화와 분포 추정 등의 분야와 밀접한 관련이 있다.

① 군집화(Clustering): 데이터가 쭉 뿌려져 있을 때 레이블이 없다고 해도 데이터간 거리에 따라 대충 두 세개의 군집으로 나눌 수 있다. 이렇게 x만 가지고 군집을 학습하는 것이 군집화이다.

② 분포 추정(Underlying Probability Density Estimation): 군집화에서 더 나아가서, 데이터들이 쭉 뿌려져 있을 때 어떤 확률 분포에서 나온 샘플들인지 추정하는 문제이다.

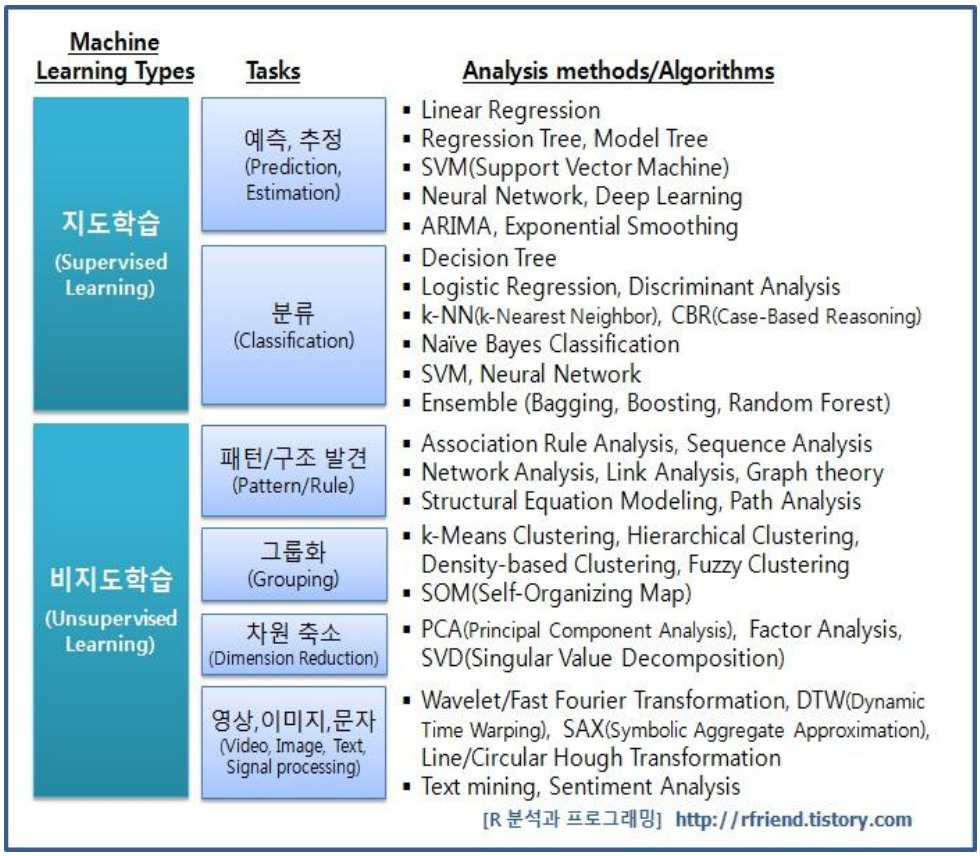
2.3 반지도 학습 (Semisupervised learning)

레이블이 있는 데이터와 없는 데이터 모두를 활용해서 학습하는 것인데, 대개의 경우는 다수의

레이블 없는 데이터를 약간의 레이블 있는 데이터로 보충해서 학습하는 종류의 문제를 다룬다.

2.4 강화 학습(Reinforcement Learning)

위의 두 문제의 분류는 지도의 여부에 따른 것이었는데, 강화학습은 조금 다르다. 강화학습은 현재의 상태(State)에서 어떤 행동(Action)을 취하는 것이 최적인지를 학습하는 것이다. 행동을 취할 때마다 외부 환경에서 보상(Reward)이 주어지는데, 이러한 보상을 최대화 하는 방향으로 학습이 진행된다. 그리고 이러한 보상은 행동을 취한 즉시 주어지지 않을 수도 있다(지연된 보상). 이 때문에 문제의 난이도가 앞의 두개에 비해 대폭 상승하며, 시스템을 제대로 보상하는 것과 관련된 신뢰 할당 문제 라는 난제가 여전히 연구원들을 괴롭히고 있다. 대표적으로 게임 인공지능을 만드는 것을 생각해볼 수 있다.



[그림 1] 기계학습 형태

[Fig. 1] Machine Learning Types

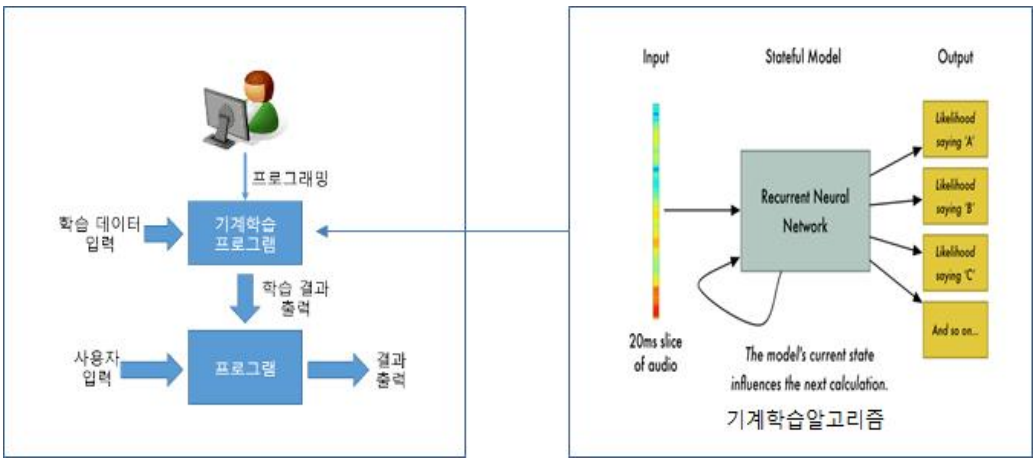
ⓒ 2018 NCISS

체스에서 현재 나와 적의 말의 배치가 State가 되고 여기서 어떤 말을 어떻게 움직일지가 Action이 된다. 상대 말을 잡게 되면 보상이 주어지는데, 상대 말이 멀리 떨어져 이동할 때 까지의 시간이 필요할 수 있으므로, 상대 말을 잡는 보상은 당장 주어지지 않는 경우도 생길 수 있다. 심지어는 그 말을 잡은 것이 전술적으로는 이익이지만 판세로는 불이익이라 다 끝났을 때 게임을 질 수도 있다. (지연된 보상). 따라서 강화학습에서는 당장의 보상값이 조금은 적더라도, 나중에 얻을 값을 포함한 보상값의 총 합이 최대화되도록 Action을 선택해야 하며, 게다가 행동하는 플레이어는 어떤 행동을 해야 저 보상값의 합이 최대화되는지 모르기 때문에, 미래를 고려하면서 가장 좋은 선택이 뭔지 Action을 여러 방식으로 수행하며 고민해야 한다. 좋은 선택이 뭔지 Action을 찾는 것을 탐색, 지금까지 나온 지식을 기반으로 가장 좋은 Action을 찾아 그것을 수행하는 것을 활용한다고 하여, 강화학습을 푸는 알고리즘은 이 둘 사이의 균형을 어떻게 잡아야 할지에 초점을 맞춘다. 위 방법들과는 다르게 실시간으로 학습을 진행하는 게 일반적이다[6][7]. [그림 1]에 기계학습 형태를 나타내었다.

3. 고장진단시스템 설계

본 논문은 소음을 이용한 기계학습 기반 자동차 고장진단 시스템 및 방법에 관한 것으로, 더 상세하게는 센서에 의해 검출된 자동차의 하부 및 엔진 룸 소음을 클라우드 서버에 전송하여 기저장된 자동차 소음 패턴 데이터를 기계학습하여 자동차의 고장을 진단할 수 있는 소음을 이용한 기계학습 기반 자동차 고장진단 시스템 및 방법에 관한 것이다. 본 논문에서 해결하고자 하는 기술적 과제는, 자동차에서 발생되는 하체 및 엔진룸의 소음을 다각도로 분석하여 위험의 정도를 다단계로 나누어 운전자에게 인식시키며, 고장이라고 진단된 부품의 수리의 지급도를 안내하여, 자동차 고장에 의한 교통사고 발생을 방지할 수 있는 소음을 이용한 딥러닝 기반 자동차 고장진단 시스템 및 방법을 제공함에 있다. 또한 본 발명이 해결하고자 하는 다른 기술적 과제는, 운전자의 위치 정보를 수집하여, 고장 부품의 수리 지급도에 따라 주변의 정비소를 자동으로 안내하여, 수리를 받을 수 있도록 하는 소음을 이용한 딥러닝 기반 자동차 고장진단 시스템 및 방법을 제공함에 있다. 그리고 진단서버에서 진단한 고장 부품과 정비소단말을 통해 피드백되는 실제 고장 부품의 정보를 확인하여, 진단의 정확도를 판단할 수 있으며, 새로운 소음 데이터 패턴을 저장하고 그에 대한 고장 부품을 매칭하는 방식을 사용하여 고장 진단의 정확도를 향상시킬 수 있는 소음을 이용한 딥러닝 기반 자동차 고장진단 시스템 및 방법을 제공함에 있다. 기계학습의 개략적인 알고리즘은, 회귀, 식별, 군집화, 확률 분포 추정, 마르코프 연쇄, 마르코프 무작위장, 은닉 마르코프 모델 등을 기반으로 설계 및 개발 된다. 아래 [그림 2]에 시스템 구성도를 나타내었다[8][9].

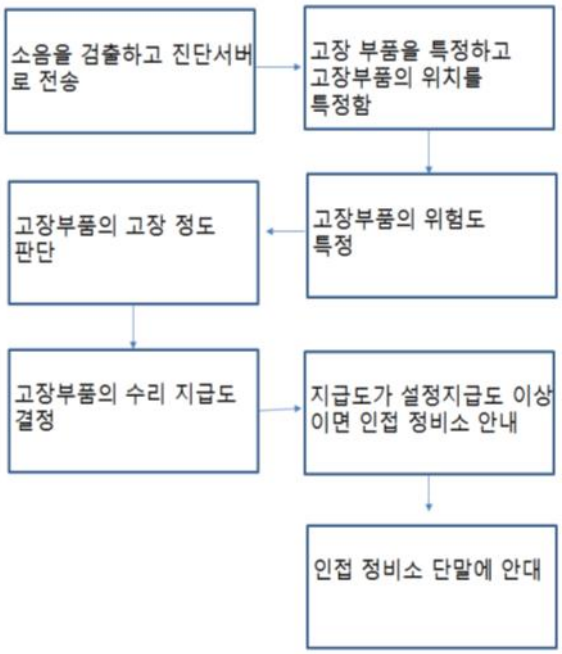
ⓒ 2018 NCISS



[그림 2]시스템 구성

[Fig. 2] System Configuration

아래 [그림 3]은 소음을 이용한 딥러닝 기반 자동차 고장진단 방법의 순서도이다.



[그림 3]시스템 흐름도

[Fig. 3] System Flow-Chat

소음을 이용한 딥러닝 기반 자동차 고장진단 방법은, 사용자단말에서 다수의 음향 센서로부터 주행 중 자동차의 소음데이터를 수집하고, 수집된 소음데이터를 진단서버로 전송하는 단계와, 진단

ⓒ 2018 NCISS

서버에서 소음데이터를 기저장된 소음데이터들과 비교하여 고장 부품을 진단하되, 다수의 음향 센서 에서 검출된 소음의 강도에 따라 고장 부품의 위치를 특정 하는 단계와, 특정된 상기 고장 부품의 위험도를 기저장된 부품별 위험도 정보에서 추출하는 단계와, 소음데이터의 크기와 발생 주기를 이용하여 고장 부품의 고장 정도를 판단하는 단계와, 상기 고장 부품의 위험도와 고장 정도를 고려하여 수리 지급도를 선택하는 단계와, 상기 고장 부품, 위험도, 고장 정도 및 수리 지급도를 상기 사용자단말로 전송하여 표시되도록 하는 단계와, 상기 단계의 수리 지급도가 설정 지급도 이상인 경우에 상기 사용자단말의 위치와 인접한 정비소를 안내하는 단계와, 상기 단계가 수행됨을 상기 인접한 정비소의 정비소단말로 전송하는 단계를 포함한다[10]. 아래 [그림 3]에 주요 시스템 흐름도를 나타 내었다.

4. 결론

본 논문에서는 기계학습을 이용한 자동차 고장진단 시스템 설계에 관한 것으로, 설치위치에서 주행 중인 자동차의 소음을 검출하는 다수의 음향센서를 포함하여 부품 이상에 따라 발생되는 소음데이터를 검출하여 전송하는 소음검출부와, 상기 소음검출부에서 검출된 소음데이터를 외부로 전송하며, 수신된 정보를 표시하는 사용자 단말과, 상기 소음데이터를 분석하여 고장 부품의 종류와 위치를 특정하고, 고장 부품의 종류에 따라 위험도를 결정하고, 소음데이터의 크기와 발생빈도를 이용하여 고장 정도를 판단하여 수리 지급도를 결정하여, 상기 고장 부품의 종류, 위치, 위험도, 수리 지급도 정보를 상기 사용자단말로 전송하는 진단 서버를 포함 한다. 이를 통해서 기계학습 관련 분야 산업 발전에 도움이 되었으면 한다.

ISSN: 2384-101X JNCIST
ⓒ 2018 NCISS

References

[1] Bae Jae-Hwan, "Vehicle Fault Diagnosis System and Method Using Noise" Korea Patent Office (KR), Application No. 1020170119733, 18 September (2017).

[2] Provost, F., and R. Kohavi. "Glossary of terms." Journal of Machine Learning 30.2-3 (1998): 271-274.

[3] Ron Kohavi, Foster Provost, "Glossary of terms", Machine Learning. 30, (1998): 271 – 274.

[4] Nasrabadi, Nasser M. "Pattern recognition and machine learning." Journal of electronic imaging 16.4 (2007): 049901.

[5] Friedman, Jerome H. "Data Mining and Statistics: What's the connection?." Computing Science and Statistics 29.1 (1998): 3-9.

[6] Dickson, Ben. "Exploiting machine learning in cybersecurity." TechCrunch. Retrieved (2017): 05-23.

[7] Wernick, Miles N., et al. "Machine learning in medical imaging." IEEE signal processing magazine 27.4 (2010): 25-38.

[8] Mannila, Heikki. "Data mining: machine learning, statistics, and databases." ssdbm. IEEE, (1996).

[9] Forni, A., and R. Meulen. "Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage." (2016).

[10] Mitchell, Tom M., and Machine Learning. "Mcgraw-hill science." Engineering/Math 1 (1997): 27.

ⓒ 2018 NCISS