효과적인 이상 진단을 위한 클러스터링의 타당성 연구

이현용*, 김낙우*, 이준기*, 이병탁*
*한국전자통신연구원
{hyunyonglee, nwkim, jungi, bytelee}@etri.re.kr

A Feasibility Study on Clustering for Effective Anomaly Detection

HyunYong Lee*, Nac-Woo Kim*, Jun-Gi Lee*, and Byung-Tak Lee*
*Electronics and Telecommunications Research Institute (ETRI)

요 약

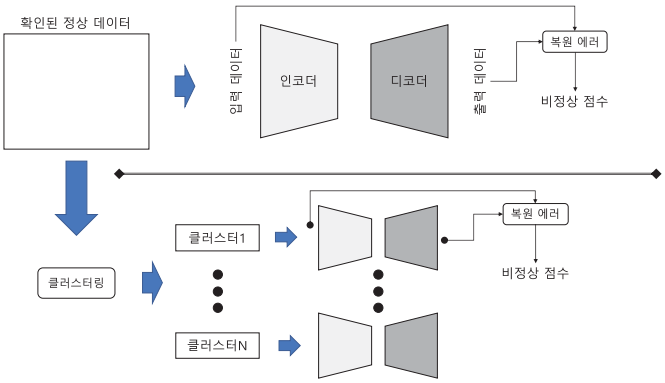
이상 진단은 주어진 데이터의 정상 유무를 진단하는 방법으로써 다양한 분야에 걸쳐 요구되는 기능이다. 이상 진단은 대상 환경에서 발생하는 데이터의 특성 등에 따라 다양한 방법으로 구현이 될 수 있는데, 본 연구에서는 정상 데이터가 다수의 클래스로 구분될 수 있는 상황에서의 이상 진단을 효과적으로 할 수 있는 방법에 대해서 다루고자 한다. 특히, 실험을 통해 정상 데이터를 유사한 데이터들끼리 구분하여 처리하는 경우와 그렇지 않은 경우의 비교를 통해서, 정상 데이터를 유사한 데이터들끼리 구분하여 이상 진단을 진행하는 방법의 타당성을 검증한다.

1. 서론

이상 진단은 기계학습 등과 같은 다양한 방법에 기반하여 주어진 데이터 (또는 상태)가 정상인지 비정상인지를 판단하는 것을 일컫는다 [1]. 이상 진단은 공장 설비 등의 이상 진단, 전력 소비의 이상 진단, 사람의 건강 상태의 이상 진단 등에 널리 요구되는 기술이다. 이상 진단은 진단 대상이 되는 환경의 데이터 특성 등에 따라서 다양한 방법으로 구현이 될 수 있는데, 주요한 연구 주제 중 하나는 효과적인 비정상 점수(abnormality score)를 찾는 것이다. 비정상 점수는 주어진 데이터의 비정상 유무 판단을 위해서 사용되며, 효과적인 비정상 점수는 정상 데이터와 비정상 데이터를 정확하게 구분할 수 있어야 한다.

본 논문에서는 보다 효과적인 비정상 점수 도출을 위한 방법을 제안하는데, 다음과 같은 상황을 고려한다. 정상 데이터는 다수의 클래스들로 구분될 수 있지만, 명시적으로 주어진 클래스 레이블은 없다. 본 논문에서 제안하는 방법의 핵심은, 정상 데이터를 구분없이 하나의 모델로 처리하는 것과 레이블은 없지만 클러스터링을 통해서 임의의 클래스로 분류한 뒤 클래스 별로 모델을 만들어 처리하는 것에는, 도출되는 비정상 점수 간의 유용성 측면에서 차이가 있을 것이라는 점이다. 본 논문에서는 본 연구의 시작점으로써 실험을 통해 위의 두 경우에서의 비정상 점수 간의 유용성을 비교한다.

2. 클러스터링 기반 이상 진단

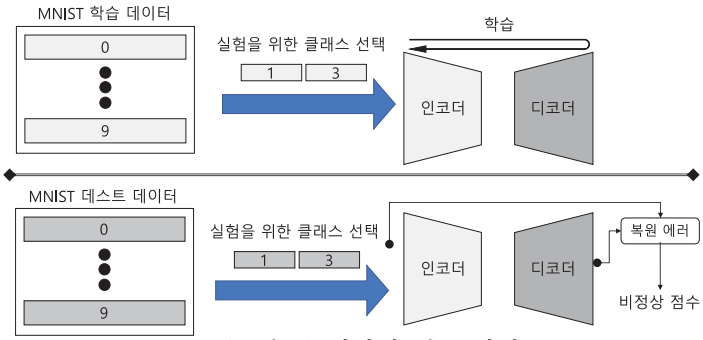


(그림 1) 클러스터링 기반 이상 진단 구조.

본 연구에서 추구하는 이상 진단 구조는 그림 1 에 표현되어있다. 확인된 정상 데이터가 주어졌다고 보고, 주어진 정상 데이터에 기반하여 이상 진단을 위한 모델을 어떻게 구성하느냐가 관건이다. 종래의 대부분의 방법은, 정상 데이터의 구분없이 전체 정상 데이터를 기반하여 하나의 모델을 학습하고, 이를 기반으로 이상 진단을 진행한다. 예를 들어, 주어진 정상 데이터에 기반하여 하나의 오토인코더 모델[2]을 만들 수 있고, 복원 에러를 비정상 점수로 사용하여 이상 진단을 진행할 수 있다. 이 때, 복원 에러가 지정된 기준 이상인 경우에, 비정상으로 간주할 수 있다. 반면, 본 연구에서 추구하는 방법은, 주어진 정상

데이터 전체를 위한 하나의 모델을 만들기 보다, 정상 데이터를 유사한 특성을 지닌 클러스터로 분류한 뒤 클러스터 별로 모델을 만드는 것이다. 정상 데이터의 클러스터링을 위해서 K-means 클러스터링[3]과 같이 잘 알려진 방법을 사용할 수 있다. 단, 몇 개의 클러스터로 구분해야 하는지는 본 논문의 관심 사항이 아니다. 클러스터링 기법을 통해 주어진 정상 데이터가 다수의 클러스터로 구분된 후에, 클러스터 별로 하나의 모델을 만들 수 있다. 예를 들어, 클러스터 별로 하나의 오토인코더 모델을 생성할 수 있다. 이 경우, 주어진 테스트 데이터의 이상 진단을 위해서는, 클러스터링 기법을 통해 테스트 데이터가 어느 클러스터에 속하는지는 먼저 판별한 후에, 해당 클러스터의 오토인코더 모델을 적용하여 복원 에러를 추출하고 이를 기반으로 이상 진단을 진행할 수 있다.

전술한 방법과 같이 주어진 정상 데이터를 클러스터링을 통해 다수의 클러스터로 구분하는 것은, 클러스터 별 모델을 해당 클러스터의 데이터에 특화시킴으로써 비정상 데이터의 경우 비정상 점수가 더 극대화되도록하기 위함이다. 다시 말하면, 상이한 특성을 보이는 정상 데이터를 하나의 모델을 통해 처리하는 것보다, 유사한 특성을 보이는 정상 데이터들끼리만 묶어서 하나의 모델을 통해 처리하는 방법의 경우에 해당 모델은 해당 클러스터의 데이터 특성을 더 잘 이해하고 표현하게 되어서, 비정상 데이터의 경우 복원 에러가 더 극대화되게 된다.



(그림 2) 타당성 검증 방법.

3. 실험 기반 타당성 검증

2 장에서 전술한 클러스터링 기반 이상 진단 기법의 타당성을 검증하기 위해서 기초 실험을 진행한다. 그림 2 는 이러한 과정을 보여준다. 실험의 목적은, 클러스터 별로 포함된 클래스의 수에 따른 그리고 클러스터에 포함된 클래스 간의 유사성 정도에 따른 비정상 점수 유용성 비교이다. 실험을 위해서 10 개의 클래스로 구성되는 MNIST 데이터[4]를 사용하였다. MNIST 는 0 부터 9 까지의 손글씨에 대한 데이터이다. 임의의 수의 클래스를 하나의 클러스터로 간주하고, 해당 클러스터의 학습 데이터에 기반하여 하나의 오토인코더 모델을 학습하고, 테스트 데이터를 적용하여 복원 에러를 점검한다. 비정상 점수로 사용되는 복원 에러는 오토인코더의 입력 데이터와 출력 데이터 간의 mean squared error 를 사용한다. 비정상 점수로 사용되는 복원 에러의 값이 작을수록 비정상 점수의 유용성이 더 크다고 볼 수 있다. 이는 해당 클러스터에 속한 데이터들을 더 잘 표현한다고 볼 수 있기 때문이다. 성능 검증 목적을 위해 별도의 클러스터링은 진행하지 않고, 공개된 클래스 정보를 클러스터 정보로 사용하였다. 다양한 클러스터 경우의 성능 비교를 위해서, 클러스터에 속하는 클래스 수를 상이하게 실험을 진행하였다. 실험은 Tensorflow 2.0 버전에 기반하여 진행하였다.

<표 1> 타당성 실험 결과

클러스터에 속한 클래스들	복원 에러 평균	복원 에러 표준편차
0	0.0141	0.005
1	0.004	0.0034
2	0.0172	0.0057
3	0.0152	0.0057
4	0.013	0.0049
5	0.0164	0.0055
6	0.0131	0.0054
7	0.0104	0.0054
8	0.0179	0.0064
9	0.0115	0.0056
1,3	0.01	0.0074
2,4	0.0166	0.0058
3,4	0.0156	0.0057
5,6	0.0157	0.0058
1,7,9	0.009	0.0058
0,3,4,7,8	0.0164	0.0063
1,2,5,6,9	0.0144	0.0075
0,1,2,3,4,5,6,7,8,9	0.016	0.0072

표 1 은 타당성 실험 결과를 보여준다. 가장 왼쪽의 열은 한 클러스터에 속한 클래스들을 보여준다. 가운데 열은 클러스터에 속한 모든 테스트 데이터에 대한 복원 에러의 평균을, 맨 오른쪽 열은 복원 에러의 표준 편차를 보여준다. 전술하였듯이, 복원 에러 평균이 낮을 수록 비정상 점수의 유용성이 높다고 판단할 수 있다. 0 부터 9 까지의 모든 클래스를 하나의 모델로 처리한 경우, 복원 에러의 평균은 0.016 이다. 반면, 하나의 클래스별로 모델을 구성한 경우에는 2,5,8 클래스의 경우를 제외하고는 더 낮은 복원 에러 평균

값을 보인다. 특히, 클래스 1 의 경우 복원 에러 평균은 0.004 이며 복원 에러 표준편차도 0.0034 로 매우 낮다. 추가로 점검해본 경우는, 2,4 와 같이 서로 상이한 숫자 모양을 보이는 클래스들을 하나의 클러스터로 묶은 경우와 1,7,9 처럼 비슷한 숫자 모양을 보이는 클래스들을 하나의 클러스터로 묶은 경우이다. 숫자 모양이 비슷한 경우는 비슷한 특성을 보이는 것으로 이해될 수 있다. 2,4 클래스를 묶은 경우, 모든 클래스를 하나의 클러스터로 묶은 경우와 유사한 복원 에러 평균을 보이는 반면, 1,7,9 를 묶은 경우 복원 에러 평균은 0.009 로 매주 낮아졌다. 표 1 에 보이는 나머지 경우에서도 관찰할 수 있는 것처럼, 이처럼 유사한 특성을 보이는 클래스끼리 클러스터로 묶는 것은 비정상 점수의 유용성을 향상시키는 것을 볼 수 있다. 반면, 관련이 없어나 특별한 구분없이 모든 데이터를 하나의 모델로 처리하는 경우에는 비정상 점수의 유용성이 낮아지는 것을 볼 수 있다.

4. 결론

효과적인 이상 진단을 위한 핵심 기술 중 하나는 정상 데이터와 비정상 데이터를 더 잘 구분하기 위한 비정상 점수를 도출하는 것이다. 본 논문에서는, 확보된 정상 데이터를 클러스터링 기법을 통해 다수의 클러스터로 구분하고 클러스터별로 모델을 구성함으로써 보다 효과적인 비정상 점수를 도출하는 방법의 타당성을 실험을 통하여 검증하였다. 본 연구의 연속을 위해서, ImageNet 과 같이 MNIST 보다 좀 더 복잡한 형태의 데이터에 기반하여 타당성 검증을 진행할 필요가 있다. 뿐만 아니라, 제안 방법의 핵심이 클러스터 별로 모델을 구성하는 것인데, 대상 테스트 데이터의 클러스터 뿐만 아니라 나머지 클러스터들의 모델들을 활용하여 비정상 점수를 도출하는 방법에 대한 연구도 의미가 있을 것으로 보인다.

감사의 글

이 연구는 정부의 ETRI R&D 프로그램(20ZK1140)의 재원을 받아 수행된 연구임.

참고문헌

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey" ACM Computing Surveys, vol. 41, no.3, July 2009.

[2] P. Baldi, “Autoencoders, unsupervised learning and deep architectures” International Conference on Unsupervised and Transfer Learning Workshop, Washington, USA, 2011, pp.37-50.

[3] T. Kanungo, et al., “An efficient k-means clustering algorithm: Analysis and implementation” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.7, pp.881-892, 2002.

[4] The MNIST DATABASE of handwritten digits, http://yann.lecun.com/exdb/mnist/