

기계학습을 활용한 IoT 플랫폼의 이상감지 시스템

임선열^{1*}, 최효근², 이규열², 이태훈², 유현창¹
 고려대학교 컴퓨터정보통신대학원¹, 삼성 SDS²
 e-mail: seonyeol.im@me.com*

Anomaly Detection System of IoT Platform using Machine Learning

SeonYeol Im, HyoKeun Choi, KyuYull Yi, TeaHun Lee, HeonChang Yu
 Graduate School of Computer & Information Technology, Korea University¹
 Samsung SDS²

요 약

많은 양의 데이터가 수집되는 산업분야에서의 IoT 플랫폼 활용도가 높아지면서 IoT 플랫폼의 성능과 이상감지가 중요한 요소가 되고 있다. 본 논문에서는 IoT 플랫폼의 데이터 수집 성능을 저해하지 않으면서 산업분야에 활용되는 디바이스의 이상을 감지하는 시스템을 제안한다. 제안한 시스템은 Soft Real-time 서비스를 제공하기 위해 데이터 전송주기를 고려한 Micro Batch 를 활용했으며, 실험에는 산업분야의 이상 상황에 대한 자료수집이 사전에 이루어지기 어려운 상황을 고려해 Hotelling's T^2 를 활용한 분석모델을 적용하였고 Hotelling's T^2 는 이상징후를 사전에 감지하였다.

Abstract

As the industry generates a lot of data, it is increasingly dependent on the IoT platform. For this reason, the performance and anomaly detection of IoT platform is becoming an important factor. In this paper, we propose a system model of IoT platform that detects device anomaly without performance issue. The proposed system uses Micro Batch which calculates the data transmission cycle to provide Soft Real-time service. In the industry, it was difficult to collect abnormal data, so the Hotelling's T^2 model was applied to the data analysis experiment. And the Hotelling's T^2 model successfully detected anomalies.

1. 서론

최근 IoT 서비스의 발전으로 다양한 분야에서 자동화된 디바이스를 활용한 서비스가 제공되고 있다. 자동화된 디바이스는 편리함을 제공해 줄 수 있지만, 예기치 못한 고장이나 이상 상황으로 인하여 서비스 공백이 생기거나 서비스 자체에 대한 신뢰도를 떨어뜨리는 문제가 발생할 수 있다. 산업분야에서는 센서 장비의 가격 하락과 IoT 서비스의 발전으로 고가의 장비에 대한 관리 혹은 서비스 공백을 최소화하기 위한 도구로 IoT 플랫폼을 활용하는 경우가 많아지고 있다. 하지만 디바이스의 이상 상황에 대해 감지하지 못하는 경우 오히려 서비스에 악영향을 미칠 수 있다.

이러한 환경에서 IoT 플랫폼의 역할은 단순히 디바이스의 상태를 모니터링하고, 제어하는 것에 그치지 않으며 연결된 디바이스의 이상 상태를 감지하는 것 또한 필수적인 기능이 되었다. 이상 상태 감지는 명시적인 임계치를 설정하는 방식으로도 가능하지만 이런 경우 디바이스에 대한 정보를 기반으로 한 사용자의 경험에 의존한다. 회전체 혹은 시계열 데이터 구조를 가진 디바이스의 경우에는 시간의 흐름에 따라 디바이스 별로 임계치가 달라

질 수 있기에 임계치를 명시적으로 설정하는 방법으로는 이상 감지에 한계가 따를 수 있다. 기계학습을 통한 데이터 분석이 보편화 되고 많은 양의 데이터가 수집되는 환경에서 기계학습을 활용한 이상 감지를 통해 임계치를 설정하는 방식의 이상 감지를 보완하거나 대체하는 것이 가능해졌다.

하지만 빅데이터를 다루는 기계학습은 실시간으로 수집되는 데이터를 처리하기에는 임계치를 설정하는 방식보다 상대적으로 많은 리소스를 소모하고 시계열 데이터 구조를 가지는 디바이스의 경우 필터링을 위해 데이터의 윈도우 처리가 필요할 수 있어 더 많은 리소스를 사용하게 된다. 이러한 문제점은 IoT 플랫폼에 성능 문제를 야기시킬 수 있으며 성능 문제는 결국 비용의 증가로 이어질 수 있다.

현재까지 IoT 플랫폼에서의 이상감지에 대한 연구가 수행되어 왔지만 기존 연구에서는 Supervised Learning 을 적용하여 이상 상황이 포함된 데이터를 확보하지 않은 경우 학습하기 어려운 문제점과 이상감지 판단 처리 방식에 대한 구체적인 표현이 부족하여 실시간 데이터 처리 방식에서 나타날 수 있는 성능문제에 대한 해결이 되었는지 알 수 없었다.

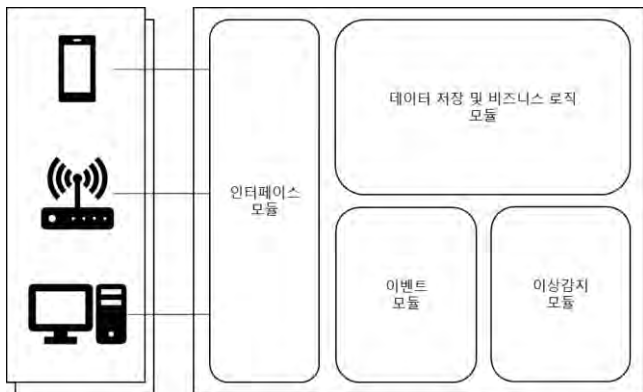
따라서 본 연구에서는 실시간 데이터 처리 방식의 성능 문제를 해결하고 Soft Real-time 서비스가 가능한 Micro Batch 처리 방식을 제안하면서 실제 서비스의 운영 환경에서 이상 혹은 장애 상황의 데이터를 수집하기 어려운 점을 고려해 Unsupervised 방식의 다변량 시계열 분석에 적합한 Hotelling's T^2 를 키오스크에 적용하여 이상 감지를 평가하였다.

2. IoT 플랫폼의 이상감지 시스템

2.1 시스템 개요

제안하는 시스템 모델은 데이터 수집과 디바이스의 관리를 목적으로 하는 IoT 플랫폼을 기반으로 한다. IoT 플랫폼의 특성상 다양한 분야의 Third-Party 에서 활용 가능하도록 범용성이 있어야 하며, 그 중 IoT 서비스의 활용도가 높은 산업분야에서는 다량의 디바이스에서 짧은 주기로 데이터를 전송하기 때문에 성능 문제를 필수적으로 고려해야 한다.

(그림 1)은 제안하는 이상감지 시스템을 적용한 IoT 플랫폼의 시스템 구성도를 표현한 것이다.



(그림 1) 이상감지 IoT 플랫폼의 시스템 구성도

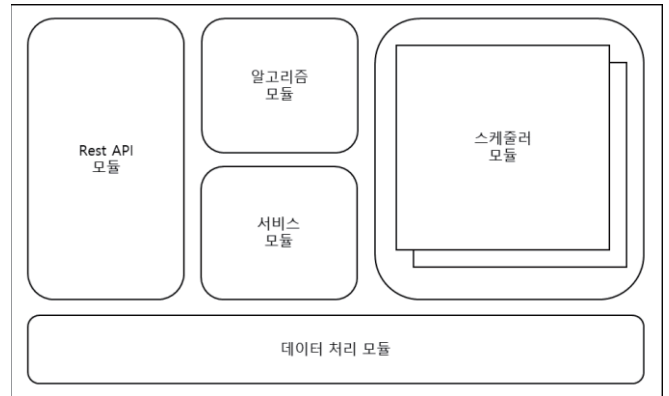
제안하는 IoT 플랫폼은 크게 네 개의 영역으로 나눌 수 있다. 먼저 다양한 디바이스와의 연결을 위해 인터페이스를 담당하는 인터페이스 모듈, 데이터 저장 및 비즈니스 로직을 처리하는 데이터 저장 모듈, 임계치 설정을 통한 이상 알림과 이벤트 처리를 수행하는 후 처리 모듈, 그리고 마지막으로 본 연구에서 제안하는 기계학습을 활용한 이상감지 모듈이다.

디바이스는 IoT 플랫폼과 등록, 초기화, 제어 등의 다양한 메시지를 주고받지만 데이터의 저장을 처리하는 메시지가 대부분을 차지하며, 데이터의 저장을 처리하는 메시지는 플랫폼 내에서 메시지 수신부터 저장, 후 처리 모듈까지 모두 높은 성능을 유지한 상태의 동기방식으로 처리가 가능한 수준이다.

하지만 향후 다양한 종류의 기계학습이 적용될 수 있는 이상감지 모듈을 기존 데이터 저장 메시지 처리에 연결하여 실시간 동기방식으로 처리하는 경우 처리 지연으로 인한 데이터 유실 혹은 서비스 장애 등의 문제가 발생할 수 있으며, 비동기 방식으로 처리한다고 해도 Sliding Window 방식을 활용한 데이터 분석을 처리하는 경우에는 기존 데이터의 호출 부분

에 있어 많은 리소스를 필요로 하는 등의 IoT 플랫폼 자체에 대한 서비스 문제를 야기시킬 수 있다.

Micro Batch 방식으로 이상감지를 처리하는 경우 디바이스 혹은 디바이스 모델 별 데이터 전송 주기를 계산하여 Batch 주기를 자동으로 설정하고 이를 통해 Real-time 에 준하는 Soft Real-time 으로 서비스 제공이 가능하며, IoT 플랫폼의 데이터 저장 성능에 영향을 미치지 않으면서 이상감지가 가능하다. (그림 2)는 이상감지 모듈의 구조를 표현한 것이다.



(그림 2) 이상감지 모듈

이상감지 모듈은 내부적으로 IoT 플랫폼에 제공할 분석 모델 업데이트 기능과 시뮬레이션 기능의 인터페이스를 처리하는 Rest API 모듈과 이상감지를 수행하는 스케줄러 모듈로 구성되어 있으며, 스케줄러 모듈은 수집되는 데이터의 전송 주기를 계산하여 계산된 주기에 의해 이상감지를 수행한다. 사전에 학습시킨 분석 모델을 사용자가 화면에서 파일로 저장하게 되며, 이상감지 모듈에서는 최초 분석 시에 데이터베이스를 통하여 분석 모델에 접근하지만 이후에는 파일로 저장 후 메모리를 활용한다.

2.2 이상감지 기계학습 모델

실제 서비스되는 운영상황 중에는 장애가 발생하면 치명적이기 때문에 장애 상황과 관련한 데이터를 충분히 수집하는 것은 어려운 일이다. 또한 IoT 플랫폼과 센서를 통해 다양한 종류의 데이터를 수집하고 있으며, 관리하는 디바이스가 많은 경우 수집되는 데이터를 복합적으로 분석하여 이상상태에 대한 단순한 알림을 우선적으로 처리하는 것이 이상감지에 적합할 수 있다. 본 실험에서는 엔진 크랭크 베어링 장비에 진동을 측정하는 센서 네 개를 장착하여 수집된 데이터를 분석하였다. 분석 기법으로는 단변량 분석은 상관성 있는 변수들이 결합하여 임의의 공정에 미치는 영향을 간과할 수 있기 때문에 통계적으로 종속변수의 상관성을 고려한 상태에서 여러 개의 종속변수를 동시에 분석할 수 있는 다변량 분석 중 Hotelling's T^2 를 적용하였다.

Hotelling's T^2 는 선형조합된 점수들의 차이를 검증하기 위해 개발된 모델이다. 단변인 독립 t 검정이 하나의 종속변인에 대한 집단 간 차이를 검증하는 것

이러면, Hotelling's T^2 는 여러 종속변인들에 대한 집단 간 차이를 검증하는 t 검증이라고 할 수 있다. 중다변인변량분석(MANOVA)에서 산출되는 많은 행렬들이 Hotelling's T^2 를 계산하는 데 사용되고, 이는 분포의 변산 정도를 알려준다.

3. 실험 및 평가

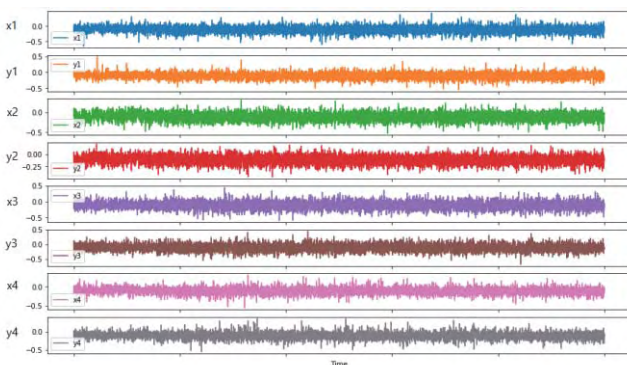
3.1 데이터 구조

실험에는 엔진 크랭크 베어링 장비의 모니터링을 위하여 베어링에 부착된 네 개의 진동 센서 데이터를 활용하였다. 데이터 수집 주기는 약 2 분이며 1 개월 간 수집된 21,561 건의 데이터를 분석에 활용했다.

Time	x1	y1	x2	y2	x3	y3	x4	y4
2017-11-15 5:08:46	-0.183	-0.02	-0.044	-0.173	-0.146	-0.171	-0.117	-0.085
2017-11-15 5:09:02	-0.173	0.039	-0.237	-0.193	-0.293	-0.046	-0.029	-0.181
2017-11-15 5:09:19	-0.083	-0.156	-0.037	-0.095	-0.132	-0.264	-0.098	-0.186
2017-11-15 5:09:36	0.012	-0.278	0.056	-0.054	-0.02	-0.115	-0.154	-0.037
2017-11-15 5:09:53	-0.037	-0.056	0.054	-0.09	-0.039	-0.081	-0.2	-0.129
2017-11-15 5:10:10	-0.1	-0.188	-0.205	-0.117	-0.098	-0.242	-0.093	-0.225
2017-11-15 5:10:27	-0.164	-0.244	-0.161	-0.183	-0.11	-0.127	0.09	-0.1
2017-11-15 5:10:44	-0.156	-0.107	-0.193	-0.229	0.012	-0.227	-0.073	-0.09
2017-11-15 5:11:01	-0.117	-0.212	-0.305	-0.125	0.022	-0.176	-0.125	-0.2
2017-11-15 5:11:18	-0.093	-0.12	-0.044	-0.178	-0.129	-0.029	-0.063	-0.09
2017-11-15 5:18:46	-0.054	-0.103	-0.039	-0.117	0.024	-0.12	-0.171	-0.02
2017-11-15 5:19:02	-0.229	-0.056	-0.229	-0.11	-0.193	-0.205	-0.161	-0.264
2017-11-15 5:19:19	-0.3	-0.063	-0.22	-0.117	-0.208	-0.125	-0.134	-0.129
2017-11-15 5:19:36	-0.349	-0.088	-0.12	-0.034	-0.317	-0.203	0.034	-0.146
2017-11-15 5:19:53	0.056	-0.203	-0.129	-0.02	-0.154	-0.037	-0.063	-0.188
2017-11-15 5:20:10	-0.076	-0.227	-0.115	-0.11	-0.095	-0.073	-0.286	-0.044
2017-11-15 5:20:27	-0.051	-0.2	-0.029	-0.034	-0.242	-0.164	-0.049	-0.122
2017-11-15 5:20:44	-0.156	-0.247	-0.205	-0.154	-0.139	-0.308	-0.034	0

(그림 4) 평가를 위해 사용된 데이터 구조

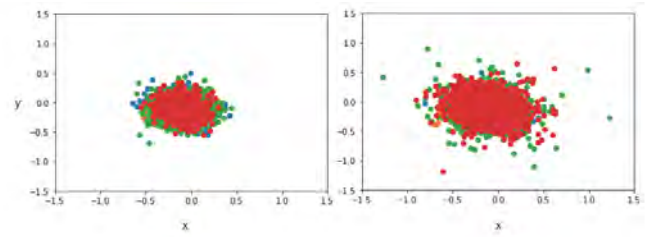
(그림 4)의 데이터 구조는 장비에 부착된 네 개의 진동 센서에 대한 일부 값이다. 센서는 진동을 x 축과 y 축의 두가지 속성으로 구분하여 전송하며 이러한 센서 네 개가 하나의 장비에 부착되어 있다. 엔진 크랭크 베어링은 사이클이 있는 장비로 일정한 형태로 데이터를 전송하는 것이 일반적인 형태이며, 정상 범주의 데이터 형태는 (그림 5)와 같다.



(그림 5) 정상 범주의 데이터 그래프

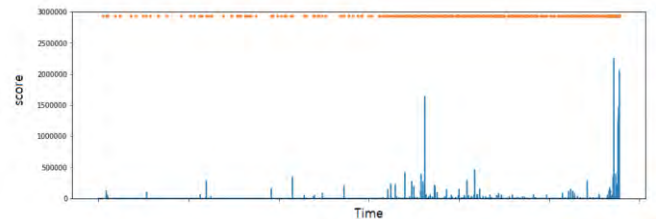
3.2 실험 결과 및 평가

실험에 활용된 Hotelling's T^2 통계량은 하나의 관측값이 다변량 정규분포의 중심에서 얼마나 떨어져 있는가를 측정하는 값으로 정상 데이터와 비정상 데이터의 분포를 그래프로 확인하면 (그림 6)과 같다.



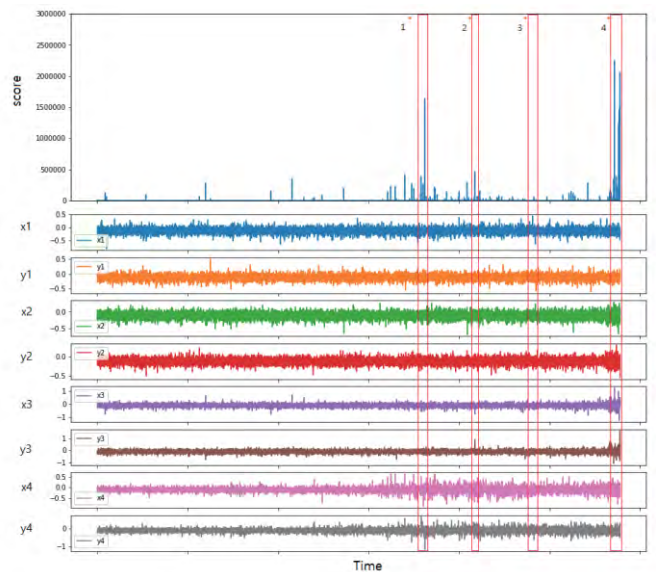
(그림 6) 정상 데이터와 이상 데이터의 분포

정상 범주의 데이터는 x, y 속성이 -0.5 와 0.5 사이에 비교적 고르게 분포되어 있는 반면 비정상 범주의 데이터는 고르게 분포되어 있는 범위가 -1.0 과 1.0 으로 다소 확장됐으며 해당 범위도 벗어나는 데이터가 많았다. 모델링에는 정상 범주의 데이터를 활용하였다.



(그림 7) 필터가 적용되지 않은 예측 결과

정상 범주의 데이터를 학습 후 이상 데이터가 포함된 테스트 데이터를 활용하여 예측했을 때 다소 높은 비율로 이상을 판단하는 결과가 나타났다. (그림 7)의 그래프에서 막대 그래프는 T^2 의 score 를 나타내고, 알람은 그래프 상단에 점으로 표시되었다. T^2 에 민감하게 반응하여 일시적인 변화로 인한 과도한 알람을 줄이기 위해 Poisson filter 를 적용하였고, 필터링까지 거친 후의 예측 결과는 (그림 8)과 같다.



(그림 8) 필터가 적용된 예측 결과

(그림 8)의 첫 번째 그래프는 (그림 7)과 같이 score 와 alarm 을 나타내며, 하단의 x1 에서 y4 의 그래프는 예측에 활용된 실제 데이터를 나타낸다. 실제

데이터에서 이상 상황이 발생한 시점은 4 로 표기된 마지막 영역이며 시각은 23:31:37 이며 그에 앞서 alarm 이 발생한 시각은 동일한 날의 14:39:30 으로 약 9 시간 앞서 이상을 감지할 수 있었다.



(그림 9) 구현된 시스템에서의 이상 감지 화면

(그림 9)는 제안하는 시스템 모델을 적용하여 실현한 삼성 SDS 사의 Brightics IoT 이상 감지 모니터링 화면이다. 이상 상황을 시스템에서 재현하기 위해 전송 주기를 200ms 로 설정하여 수집된 데이터를 시스템에 전송했고 기존 플랫폼의 데이터 수집 성능에 영향을 미치지 않으면서 그림 9 와 같이 이상을 사전에 감지하는 것을 확인할 수 있었다. 그림 9 의 첫번째 그래프는 T^2 의 score 와 alarm 을 나타내는 그래프이고, 두번째 그래프는 이상 상태의 Root Cause 를 확인할 수 있는 그래프이다. 하단의 그래프는 속성별 실제 데이터와 이상 기간을 나타낸 그래프로 사용자에게 이상 상황과 관련한 다양한 정보를 제공한다.

4. 결론

본 논문은 사물인터넷 환경을 통해 연결된 디바이스의 데이터를 기준으로 이상 장애를 탐지하기 위한 머신러닝 기반 이상 감지 시스템을 제안 및 구현하였다. 제안된 시스템은 디바이스 모델 별 데이터 전송 주기를 계산하여 동작하는 Micro Batch 를 통해 IoT

플랫폼의 데이터 수집 성능을 저해하지 않으면서 Soft Real-time 서비스를 제공할 수 있었다. 이상 감지를 위한 머신러닝 실험에는 Hotelling's T^2 를 적용하여 Unsupervised Learning 방식으로 이상 상황이 발생하기 수 시간 전에 감지해내는 것을 확인할 수 있었고 Poisson filter 를 적용해 False Alarm 을 최소화할 수 있었다. 이러한 결과를 통해 이상 상황에 대한 데이터를 사전에 수집하기 어려운 다양한 산업 분야에도 IoT 플랫폼을 통한 머신러닝 기반 이상감지를 적용할 수 있을 것으로 기대한다.

제한하는 시스템의 한계는 Data Scientist 를 통해 분석모형을 생성해야 한다는 점이 있다. 이러한 한계는 더 많은 산업분야에 IoT 플랫폼이 활용되는데 사용성 면에서 제한이 될 수 있으며, 향후 연구에서는 이러한 한계를 개선하기 위해 Data Scientist 의 개입 없이도 이상 감지가 가능하도록 단변량 데이터분석을 IoT 플랫폼에 도입하고자 한다.

참고문헌

- [1] J. H. Kang, S. B. Kim, "A clustering algorithm-based control chart for inhomogeneously distributed TFT-LCD processes", International Journal of Production Research, 2013 - Taylor & Francis
- [2] S. I. Na, H. J. Kim, "Design of Anomaly Detection System Based on Big Data in Internet of Things", Journal of Digital Contents Society, Vol. 19, No. 2, pp. 377-383, Feb. 2018
- [3] S. B. Kim, J. H. Na, Y. S. Seo, "Prediction of Rainfall-Induced Slope Failure Using Hotelling's T-Square Statistic", The Journal of Engineering Geology, Vol.25, No.3, September, 2015, pp. 331-337
- [4] K. L. Jae, W. Ko, T. O. Kim, D. I. Shin, "Fault Detection, Monitoring and Diagnosis System of CNG Stations", The Korean Institute of Gas, 2012.5, pp. 68-71
- [5] J. M. Park, B. W. Hwang, W. D. Cho, K. S. Kim, K. H. Kim, "An Example and Application of Statistical Anomaly Detection using Elasticsearch", Proceedings of Symposium of the Korean Institute of communications and Information Sciences, 2017.1, 1232-1233
- [6] J. C. Choi, I. K. Jang, D. H. Lee, M. E. Kim, and Y. S. Soen, "Technology status and prospects for develop intelligence internet of Things," IITP Series, No.1800, 2017. [Internet]. Available : <http://www.itfind.or.kr/WZIN/jugidong/1800/file6216979542456700532-1800.pdf>
- [7] W. D. Cho, J. H. Seong, S. T. Choi, J. H. Kim, J. H. Ha, J. S. Baek, and J. Y. Lee, "Adaptive sensing and monitoring technologies for detecting big-data based on IoT multi-sensors," Journal of The Korea Institute of Information Scientists and Engineers, vol. 35, no. 7, pp. 26-34, 2017.