

RRCF 알고리즘을 활용한 RAN 장비 이상 검출에 관한 연구

이택현* · 국광호

서울과학기술대학교

A Study on RAN Equipment Anomaly Detection Using RRCF Algorithm

Taek-Hyun Lee* · Kwang-Ho Kook

Seoul National University of Science and Tech

E-mail : futp@naver.com / khkook@seoultech.ac.kr

요 약

코로나19의 팬데믹 현상으로 인하여 모바일 서비스의 활용 비중이 높아지고 있다. 그러나, 대부분의 모바일 장비에 대한 이상 현상을 장비의 알람 중심으로 인지하므로, 복잡한 장애가 발생할 경우에 직관적으로 장비의 문제 판별하기 어려운 한계가 발생한다. 이를 보완하기 위해서 본 연구에서는 장비의 알람과 성능 정보를 조합하여 직관적으로 문제를 인지할 수 있도록 RRCF 알고리즘을 활용하여 Anomaly Score 생성하였으며, 과거 장애 이력을 97% 검출하는 효과를 검증하였다.

ABSTRACT

Due to the pandemic of Corona 19, the use of mobile services is increasing. However, since anomalies in most mobile devices are recognized by the device's alarm, it is difficult to intuitively determine the problem of the device when a complex failure occurs. To compensate for this, in this study, the Anomaly Score was created by RRCF algorithm to intuitively recognize the problem by combining the alarm and performance information of the equipment, and the effect of detecting 97% of the past failure history was verified.

키워드

Radio Access Network, Anomaly Detection, Fault Management, Roubust Random Cut Forest

I. 서 론

코로나 19로 인하여 이동통신 서비스를 활용하는 언택트 기술이 발전하고 있다. 다수의 사람이 이동통신 서비스를 활용하여 음식 배달, 식료품 배달, 콘텐츠 감상 등 일상생활을 하고 있다. 국내의 경우에 팬데믹 이후 신규 모바일 앱 다운로드 비중이 음식 배달 36%, 식료품 배달 34%, 비디오 콘텐츠 33% 등으로 증가하였으며[1], 중국의 경우에 모바일 트래픽이 약 33.3%가량 증가하였다. 이러한 변화를 살펴볼 때 향후 이동통신 서비스 활용 비중이 지속하여 증가할 것으로 예상된다[6].

이동통신 서비스는 코어망(Core Network)와 무선 접속망(Radio Access Network)으로 구분된다. 코어

망은 중앙에서 전체 서비스를 총괄하는 구간이며, 무선 접속망은 가입자의 접점에서 전국적으로 서비스를 제공하는 구간이다[6]. 무선 접속망은 지역 환경에 따라서 장애 발생 빈도가 높으므로 고객 체감 품질 관리가 필요한 대상이다.

장비의 이상을 검출하기 위한 대표적인 방법으로 장비에서 생성하는 알람과 성능 정보를 활용하고 있다. 현재의 장애 감시 체계는 알람과 성능 정보를 개별적으로 확인하므로, 장비의 이상 유무를 직관적으로 판별하기 어려운 한계가 있다. 본 연구에서는 도메인 전문가가 운영 장비의 이상 유무를 직관적으로 판별할 수 있도록 장비에 대한 단변량 Anomaly Score 생성 방법을 제안한다.

II. Anomaly Score 생성 연구

* speaker

일반적으로 알려진 장비 이상 검출 방법에는 알람 기반의 검출과 성능 지표 기반의 검출로 구분할 수 있다. 알람 기반의 감시는 사후 대응 방식으로, 사전에 장애를 예방하기 어려운 문제가 존재한다. 이를 보완하기 위해서 장비의 성능 데이터를 분석하여 장비의 이상 징후를 분석하고 있으나, 복합적인 장애 상황에서 매우 많은 지표를 분석하는 것은 물리적으로 어려움이 존재한다.

이를 보완하기 위한 방안을 Fournier-Viger는 장애와 관련된 알람을 압축하여 장애를 직관적으로 인지할 수 있는 장애 알람 패턴 발굴 연구를 진행하였다[4]. Çetin은 다변량 변수를 Monte Carlo 기법으로 샘플링하고, σ 규칙을 적용하여 Anomaly Score를 생성하는 방법을 제안하였다. 예를 들어 3σ 로 이상 패턴이 검출될 경우에 99.7%의 값을 Anomaly Score로 정의하였다[2]. Guha는 단변량 데이터인 뉴욕 택시 승차 데이터를 활용하여 Anomaly Score를 생성하고 3σ 기법으로 데이터의 이상을 검출할 수 있는 모형을 제안하였다[5]. RRCF(Robust Random Cut Forest, 이하 RRCF)는 AWS 회사의 Sagemaker 모듈에서 스트리밍 이상 검출 모형으로 제공되고 있으며, 데이터를 샘플링하여 다수의 Tree를 생성하고 노드의 깊이가 낮을수록 Anomaly Score가 높은 것으로 정의하고 있다.

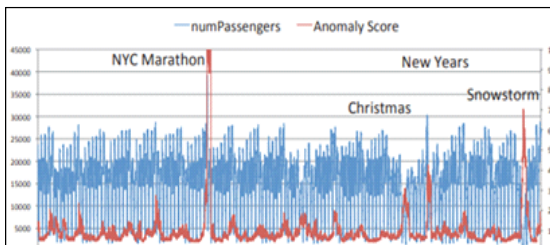


그림 1 뉴욕 택시 승차 이상탐지[5]

본 연구에서는 선행 연구를 바탕으로 AWS 회사에서 제안한 RRCF 알고리즘을 활용하여 장비에서 생성하는 다변량의 알람과 성능 지표를 단변량 Anomaly Score로 생성하고, 3σ 를 적용하여 이상을 감지하는 연구 방안을 제안하였다.

III. 장애 이상 감지 구현 방법

모바일 접속망 장비에 대한 이상탐지 모형을 개발하기 위해서, 데이터 준비, 이상검출 모델 수행, 결과 검증 절차를 진행하였다. 각 단계별 세부 수행 절차는 다음과 같다.

첫 번째, 데이터 준비 절차는 과거 장애 이력이

존재하는 무선 접속망 장비를 선별하여 알람 정보와 성능 정보를 수집하였다. 또한, 알람과 성능 정보의 발생 빈도를 분석하여, 총 107개의 특징으로 파생 변수를 생성하였으며, 장비별 수집 데이터를 1시간 통계 데이터를 결합하여 데이터마트를 생성하였다. 또한, 누락 데이터를 보완하기 위해서 데이터 보간(Interpolation) 절차를 진행하였다.

두 번째, AWS사에서 제공하는 RRCF 알고리즘을 활용하여, 다변량의 알람과 성능 정보를 결합하여 단변량의 Anomaly Score를 생성하였다.

세 번째, 생성한 Anomaly Score 정보를 3σ 로 분석하여 이상탐지 알고리즘을 수행하였다. 감시하고자 하는 도메인의 수준에 따라서 σ 값을 변경하여 이상탐지 모형을 최적화할 수 있다.

앞서 설명한 모바일 접속 장비에 대한 세부 이상탐지 수행 절차는 [그림1] 과 같다.

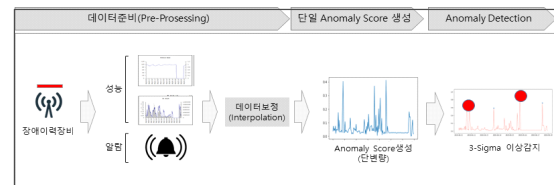


그림 2 Anomaly Detection 수행 절차

IV. 수행 결과

본 연구에서 제안하는 장비 이상탐지 모델의 정확도를 검증하기 위해서 Anomaly Detection 결과에 대하여 과거 장애 탐지 이력과 비교하는 검증을 진행하였다. 전체 장비 이상탐지 분석 대상은 장비와 시간을 결합하여 시계열 데이터 209,602개의 모수를 생성하였으며, 3σ 로 장비의 이상을 탐지한 결과 총 5,173개의 문제 후보가 탐지되었다. 이를 과거 장애 이력과 비교한 결과 97%의 탐지 정확도를 확인하였으나, 실제 과거 장애 이력 보다 과도하게 이상이 탐지되는 현상이 발생하였다. 이러한 결과는 사전에 장비의 이상을 탐지할 수 있는 정보로 활용할 수 있으나, 운영자의 과도한 업무 부하를 줄 수 있는 문제가 상존한다. 이러한 문제점을 해소하기 위해서 탐지 결과를 다시 분석하여 정확도를 개선하는 추가 연구가 필요하다.

V. 결론

본 연구에서는 모바일 접속망 장비에서 생성하는 다수의 알람과 성능 정보를 분석하여 장비의 이상을 효율적으로 검출할 수 있는 방안을 제안하였다. 이를 통하여 과거 장애 이력의 97%를 탐지할 수 있었다.

연구의 한계점으로 3σ 기법을 통하여 과거 장애 이력보다 과도하게 이상탐지가 되는 문제가 발견되었다. 이러한 결과는 운영 장비의 미세한 품질 감시를 수행하여 장애를 사전에 예방할 수 있는 정보를 제공할 수 있으나, 운영자에게 과도한 업무 부하를 발생시켜, 실제적으로 탐지 결과가 업무에 활용되지 않는 정보로 전락할 수 있다. 이를 보완하기 위해서 감시 로직을 적용하는 도메인의 감시 품질 수준에 따라서 σ 값에 대한 최적화가 필요하다. 또한, 이상 탐지 모형에 대한 탐지 정확도를 개선하기 위한 추가적인 연구가 필요하다.

References

- [1] Criteo, "2020 Report on app user behavior", 2020.
- [2] Çetin, Uzay, and Mursel Tasgin. "Anomaly Detection With Multivariate K-sigma Score Using Monte Carlo." 2020 5th International Conference on Computer Science and Engineering (UBMK). IEEE, 2020.
- [3] Erricson, "Ericsson mobility report", June 2020.
- [4] Fournier-Viger, Philippe, et al. "Discovering alarm correlation rules for network fault management." 2020.
- [5] Guha, Sudipto, et al. "Robust random cut forest based anomaly detection on streams." International conference on machine learning. PMLR, 2016.
- [6] Hwan-Guk Kim, Bo-Min Choi, Seong-Min Park, Won-Tae Shim, "Security Issues and Cyber Response Technology Considerations according to the Evolution of 5G Network Technology", Information Communication Planning and Evaluation Institute, Oct. 2019.
- [7] Mobile traffic in China in March surged 40% from a year ago due to Corona 19. Available : <https://www.yna.co.kr/view/AKR20200427142600009>.