(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년04월13일
(11) 등록번호 10-2522005
(24) 등록일자 2023년04월11일

(51) 국제특허분류(Int. Cl.)
G06F 11/30 (2006.01) *G06F 11/32* (2006.01)
G06F 9/455 (2018.01) *G06N 20/00* (2019.01)
(52) CPC특허분류
G06F 11/301 (2013.01)
G06F 11/3065 (2013.01)
(21) 출원번호 10-2021-0018674
(22) 출원일자 2021년02월09일
심사청구일자 2021년02월09일
(65) 공개번호 10-2022-0114986
(43) 공개일자 2022년08월17일
(56) 선행기술조사문헌
Jibum Hong 외 3명, "A Machine Learning based SLA-Aware VNF Anomaly Detection Method in Virtual Networks", 2020 International Conference on Information and Communication Technology Convergence(ICTC), 202
(뒷면에 계속)
전체 청구항 수 : 총 15 항

(73) 특허권자
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
홍원기
경상북도 포항시 남구 지곡로 319, 328동 304호
유재형
서울특별시 송파구 올림픽로 135, 211동 1303호
(뒷면에 계속)
(74) 대리인
특허법인이상

심사관 : 김계준

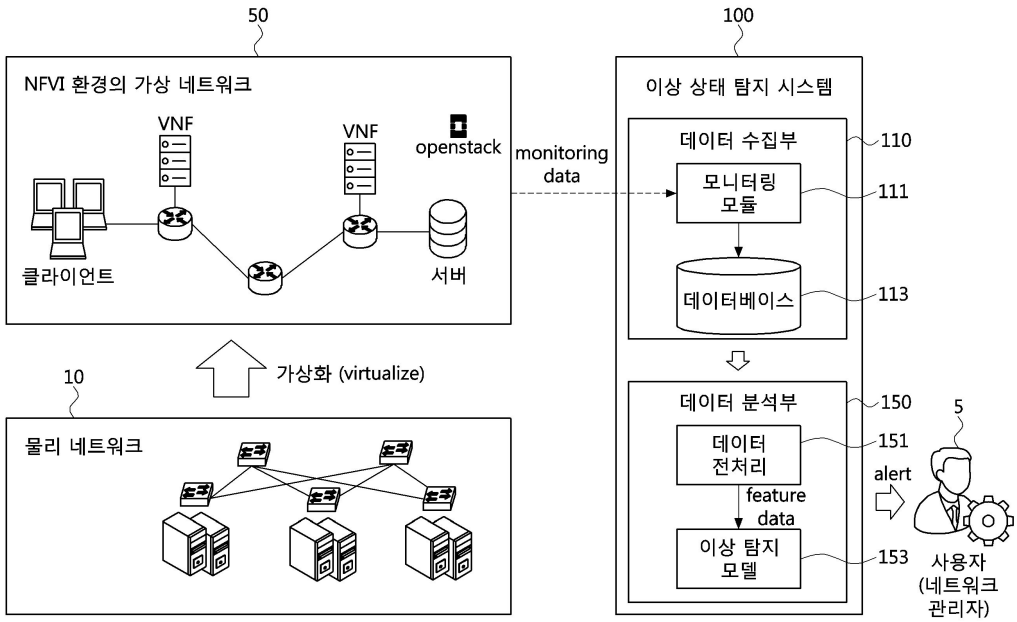
(54) 발명의 명칭 **가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법**

(57) 요 약

본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템은, 물리 네트워크에서 가상화를 통해 구성된 NFV 환경(Network Function Virtualization Infrastructure)의 가상 네트워크에서 동작하는 VNF(Virtualized Network Function)의 이상 상태를 탐지하기 위한 이상 상태 탐지 장치에 있어서, 서비스가 정

(뒷면에 계속)

대 표 도



상적으로 제공되어 생성되는 정상 상태 데이터와 결함 주입 방법을 통해 생성되는 이상 상태 데이터를 모니터링 에이전트와 모니터링 모듈을 통해 실시간으로 수집하고, 수집된 데이터는 시계열(time-series) 데이터 베이스에 저장되고, 시계열 데이터 베이스에 저장된 모니터링 데이터가 이상 상태 여부를 판단하기 위해 데이터 분석부로 전송하는 데이터 수집부 및 데이터 수집부에서 제공받은 모니터링 데이터를 전처리를 통해 이상 상태 탐지에 필요한 특성을 추출하고, 추출된 특성 데이터를 이상 상태 탐지 모델로 보내면, 이상 상태 탐지 모델은 실시간으로 들어오는 데이터를 분석하여 이상 상태 여부를 판단하고, 이상 상태가 발생한 경우 네트워크 관리자에게 통지하는 데이터 분석부를 포함한다.

(52) CPC특허분류

G06F 11/324 (2013.01)
G06F 9/45558 (2013.01)
G06F 9/5077 (2013.01)
G06N 20/00 (2021.08)
G06F 2009/4557 (2019.08)
G06F 2009/45591 (2019.08)
G06F 2009/45595 (2019.08)

(72) 발명자

홍지범

서울특별시 도봉구 도봉로136길 28, 512동 1801호

박수현

서울특별시 마포구 월드컵북로38가길 20, 101동 1003호

(56) 선행기술조사문헌

KR1020200063343 A

홍지범, “NFV 환경 관리를 위한 머신러닝 기반의 이상 탐지 방법”, 포항공과대학교 석사학위논문, 2020.12.31.

정세연 외 4명, “머신러닝 기반의 NFV 관리를 위한 모니터링 프레임워크 구조”, KNOM Conference 2018, 2018.05.11.

US20180183682 A1*

US20200104154 A1*

JP2019509681 A*

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711102868
과제번호	2018-0-00749-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발
연구과제명	인공지능 기반 가상 네트워크 관리기술 개발
기 여 율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

물리 네트워크에서 가상화를 통해 구성된 NFV 환경(Network Function Virtualization Infrastructure)의 가상 네트워크에서 동작하는 VNF(Virtualized Network Function)의 이상 상태를 탐지하기 위한 이상 상태 탐지 장치에 있어서,

서비스가 정상적으로 제공되어 생성되는 정상 상태 데이터와 결함 주입 방법을 통해 생성되는 이상 상태 데이터를 모니터링 에이전트와 모니터링 모듈을 통해 실시간으로 수집하고, 수집된 데이터는 시계열(time-series) 데이터 베이스에 저장되고, 시계열 데이터 베이스에 저장된 모니터링 데이터가 이상 상태 여부를 판단하기 위해 데이터 분석부로 전송하는 데이터 수집부; 및

데이터 수집부에서 제공받은 모니터링 데이터를 전처리를 통해 이상 상태 탐지에 필요한 특성을 추출하고, 추출된 특성 데이터를 이상 상태 탐지 모델로 보내면, 이상 상태 탐지 모델은 실시간으로 들어오는 데이터를 분석하여 이상 상태 여부를 판단하고, 이상 상태가 발생한 경우 네트워크 관리자에게 통지하는 데이터 분석부; 를 포함하고,

상기 데이터 분석부는

모니터링 데이터를 서비스 수준 협약(SLA, Servide Level Agreement) 위반 및 서비스 요청의 실패 중 적어도 하나 이상을 기준으로 정의된 이상 상태와의 관련도에 기반하여 레이블링함으로써 전처리하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템.

청구항 2

청구항 1에 있어서, 데이터 수집부는,

가상 네트워크에서 동작하는 각 가상머신의 자원 사용 상태를 주기적으로 수집하고, 수집된 모니터링 데이터를 모니터링 모듈로 보내는 모니터링 에이전트;

데이터 베이스에 저장된 시계열 모니터링 데이터를 시각화 형태로 제공하는 대쉬보드; 를 더 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템.

청구항 3

이상 상태 탐지 모델을 학습시키기 위해 NFV 환경(Network Function Virtualization Infrastructure)을 모니터링하는 NFVI 모니터링 단계;

VNF(Virtualized Network Function)의 비정상적인 상태를 발생시키는 결함 주입(fault injection) 단계;

이전 단계에서 수집된 모니터링 데이터를 서비스 수준 협약(SLA, Servide Level Agreement) 위반 및 서비스 요청의 실패 중 적어도 하나 이상을 기준으로 정의된 이상 상태와의 관련도에 기반하여 레이블링함으로써 이상 상태 탐지 모델을 학습시키기 위한 형태로 변환하는 전처리(preprocessing) 단계; 및

이상 상태 탐지 알고리즘을 통해 이상 상태 탐지 모델을 학습시키고, 학습된 이상 상태 탐지 모델을 검증한 결과를 비교하여 이상 상태 탐지 모델을 도출하는 이상 상태 탐지 모델 학습 성능 평가 단계; 를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 4

청구항 3에 있어서, 상기 방법은,

이상 상태 탐지 모델 학습 성능 평가 단계에서 도출된 이상 상태 탐지 모델을 기반으로 이상 상태 탐지 알고리즘을 통해 다시 이상 상태 탐지 모델을 학습시키는 피드백 단계를 더 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 5

청구항 3에 있어서, NFVI 모니터링 단계는,

모니터링 에이전트(agent)가 가상 네트워크에서 동작하는 각 가상머신의 자원 사용 상태인 모니터링 측정치를 주기적으로 수집하고,

모니터링 모듈(module)이 모니터링 에이전트로부터 수집된 모니터링 측정치 데이터를 수신하고, 수집된 모니터링 측정치 데이터를 시계열 데이터 데이터 베이스에 저장하고,

대쉬보드(dashboard)가 학습을 위한 데이터셋(dataset) 형태로 변환되어 데이터 베이스에 저장된 데이터가 전처리 과정을 거치고 난 후 사용자가 원하는 시각화 형태로 제공받는 단계인,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 6

청구항 3에 있어서, 결함 주입 단계는,

결함 주입(fault injection)은 실제 운영 환경에서 발생하는 이상 상태의 발생 빈도를 제어하기 위해 사용하는 기술을 이용하여 VNF가 동작하는 가상 네트워크에서 발생 가능한 소프트웨어 및 하드웨어의 이상 상태를 결함 주입 기술을 통해 발생시키는 단계인,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 7

청구항 3에 있어서, 결함 주입 단계는,

VNF가 동작하는 VM에 이상 상태를 발생시키거나, 대량의 트래픽을 전송하여 정상 서비스를 보장할 수 없을 정도의 과부하를 유발하는 결함 주입 기술을 통해 이상 상태를 발생시키는 단계인,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 8

청구항 3에 있어서, 결함 주입 단계는,

VNF가 동작하는 VM에 CPU 부하 및 메모리 부족, 디스크 I/O 엑세스 실패, 네트워크 지연, 네트워크 패킷 손실의 직접적으로 결함을 주입하는 단계이거나,

트래픽 또는 서비스에 대한 접근(access) 및 요청(request)의 허용 범위를 초과하여 들어오는 상황을 발생시켜 패킷 처리의 지연(packet processing delay) 및 커널에 의한 패킷 드롭(packet drop)을 발생하는 단계인,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 9

청구항 3에 있어서, 전처리 단계는,

모니터링을 통해 수집된 측정값들 중 정상 및 이상 상태를 판별하는데 기준이 되는 값들을 구별하여 선정하고, 수집되는 각 측정치 중 서로 중복되거나 비슷한 특성을 지니는 항목을 제거하여, VNF의 정상 및 이상 상태를 판별하는 특성들을 추출하여 그 데이터를 모델 학습에 사용하는 특성 선택(feature selection) 단계를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 10

청구항 3에 있어서, 전처리 단계는,

추출된 특성 데이터(feature data)를 지도학습 기반의 머신러닝 알고리즘에 사용할 수 있도록 각 시점의 데이터를 정상 상태 및 이상 상태로 분류하는 데이터 레이블링(data labeling) 단계를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 11

청구항 3에 있어서, 전처리 단계는,

결함 주입으로 발생시킨 시스템 및 트래픽의 과부하로 인해 VNF 내부에서 발생하는 SLA 위반을 판단할 수 있는 정보와 서비스의 요청 상태를 기준으로 이상 상태를 정의하고,

SLA 위반 및 서비스 요청 실패가 발생하는 경우를 이상 상태로, 이상 상태 이외의 상태를 정상 상태로 레이블링하여 데이터셋을 생성하는 단계인,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 12

청구항 3에 있어서, 이상 탐지 모델 학습 성능 평가 단계는,

전처리 단계에서 생성된 레이블링 데이터셋을 통해 지도학습 기반의 XGBoost 알고리즘을 사용한 학습으로 이상 탐지 모델을 생성하는 단계를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 13

청구항 3에 있어서, 이상 탐지 모델 학습 성능 평가 단계는,

결함 주입 단계 및 전처리 단계에서 SLA 위반 정보 및 응용 서비스 제공 상태를 바탕으로 레이블링된 데이터셋을 통해 XGBoost 알고리즘 기반 학습으로 이상 탐지 모델을 생성하고, 생성된 이상 탐지 모델의 분류 정확도를 검증하고 모델 성능을 평가하는 단계를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 14

청구항 3에 있어서, 모델 학습 단계는,

이상 상태 탐지 학습을 위해 선택된 특성 목록으로 측정 시각, VNF 인스턴스명, CPU - 유휴 시간, CPU - 인터럽트 처리에 소모한 시간, CPU - nice value의 프로세스를 실행하며 소모한 시간, CPU - softirq 처리에 소모한 시간, CPU - hypervisor에 의한 CPU 대기 시간, CPU - kernel 모드에서 소모한 시간, CPU - user 모드에서 소

모한 시간, CPU - I/O 대기 시간, 네트워크 인터페이스의 수신 트래픽 대역폭, 네트워크 인터페이스의 송신 트래픽 대역폭, 네트워크 인터페이스의 수신 패킷 수, 네트워크 인터페이스의 송신 패킷 수, Disk - 여유 공간, Disk - 예약된 공간, Disk - 사용 중인 공간, Disk - I/O 읽기, Disk - I/O 쓰기, Disk - I/O 수행 시간, Memory - 여유 공간, Memory - 버퍼된 공간, Memory - 캐시된 공간, Memory - 사용중인 공간, 네트워크 패킷 지연 시간을 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

청구항 15

청구항 3에 있어서, 모델 학습 단계는,

VNF 이상 탐지 모델이 사용하는 XGBoost 알고리즘의 하이퍼 파라미터 값으로 트리 개수, 트리의 최대 depth, leaf의 최소 observation 수, column 샘플링 비율, 트리당 column 샘플링 비율, early stopping에 사용할 메트릭, early stopping에 사용되는 값, L2 regularization, L1 regularization를 포함하는,

가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법에 관한 것이다.

배 경 기 술

[0003] SDN(Software-Defined Networking)/NFV(Network Function Virtualization) 기술의 급속한 발전으로 통신 사업자와 클라우드 데이터 센터 사업자들은 네트워크 기능을 가상화한 VNF(Virtualized Network Function)를 도입하여 운용하고 있으나, 점차 그 규모가 커짐에 따라 VNF의 자원 할당(resource allocation)과 성능 관리, VNF 및 VNF를 연결하는 가상 네트워크의 장애 관리(fault management) 등과 같은 새로운 관리 문제가 발생하고 있다. 이러한 SDN/NFV 전반에 걸친 관리 문제를 해결하기 위해서는 데이터 센터 내부 서버에서 동작하는 VNF가 사용하는 자원 및 가상 네트워크의 이상 상태(abnormal state)를 실시간으로 파악하고 분석해야 한다. 가상 네트워크의 자원 및 네트워크 이상 상태를 파악하기 위해 과거에는 임계값(threshold) 기반으로 이상 상태를 탐지하였다. 최근에는 머신러닝(machine learning) 기술을 접목하여 사람의 개입없이 네트워크를 관리하려는 시도들이 늘어나면서 머신러닝 기술을 기반으로 하는 이상 상태 탐지 방법도 등장하고 있다.

[0004] 하지만 기존의 임계값 기반의 탐지 방법이나 머신러닝 기반의 탐지 방법은 서버의 CPU 사용률이나 메모리 사용률과 같은 비교적 단순한 측정치(metrics)를 기준으로 이상 상태를 탐지하는 것으로서 오탐지(false alarm)를 일으킬 가능성이 크다는 문제를 가지고 있다. 본 발명에서는 서비스의 상태를 기반으로 VNF의 이상 상태를 탐지하는 방법(anomaly detection)을 제안한다. 제안하는 방법은 머신러닝 기술을 통해 VNF의 자원 및 네트워크 상태를 분석하는 방법을 포함한다.

[0005] 이상 탐지는 데이터 센터 내부에서 운용되는 물리 서버를 포함, 가상 머신 (Virtual Machine, VM) 및 VNF와 같이 NFV 환경에서 동작하는 가상 자원 및 가상 네트워크 관리와 보안의 중요한 요소이다. 네트워크 관리자는 가상화된 환경에서 제공되는 그들의 서비스들이 정상적으로 동작하고 있는지, 할당된 자원의 사용 상태는 적절한지 등을 파악하고, 상황에 맞는 정책을 실행하기 위해 이상 상태 탐지 방법을 사용한다.

[0006] 이상 탐지 방법에는 크게 시스템 자원(system resource)의 이상 상태를 탐지하는 것과 네트워크 트래픽의 이상 상태를 탐지하는 2가지 방법이 있다. 시스템 자원의 이상 상태를 탐지하는 방법은 CPU 사용량(CPU utilization), 메모리 사용량(memory usage), 디스크 I/O 엑세스(disk I/O access) 상태와 같은 측정치를 모니터링하여 CPU가 과다하게 사용되고 있거나 메모리가 부족한 상황 등을 파악하는 방법이다. 네트워크 트래픽의 이상 상태를 탐지하는 방법은 네트워크 트래픽의 평상시 정상 운용 상황을 기준으로 급격한 트래픽 증가 또는 DoS(Denial of Service)와 같은 공격 트래픽의 발생 여부를 파악하는 방법을 사용한다. 상기 두 가지 탐지 방법에 머신러닝 기술을 접목하여 이상 상태를 탐지하는 연구가 최근 많이 이루어지고 있다.

[0007] NFV 환경 관리를 위해 VNF의 이상 상태를 탐지하는 상기 2가지 방법 중 시스템 자원 기반의 탐지 방법은 과거에

는 통계적 접근 방법을 활용하여 임계값 기반으로 이상 상태를 판단하는 방법이 많이 사용되었다. 기존의 탐지 방법은 데이터 분포의 평균치에서 표준 편차의 3배가 떨어진 지점을 예외 상황으로 구분하는 3-시그마 규칙(3-sigma rule) 혹은 시계열 데이터에서 고정된 주기에 따라 변화하는 계절성 요인(seasonality factor)을 고려한 STL(Seasonal Trend decomposition using LOESS) 알고리즘 등과 같은 통계적 접근 방법을 활용하여 임계값을 설정하였다. 이러한 통계적 접근법은 이상 상태가 단일 값으로 정의될 때에는 효율적이지만, 복잡한 조건으로 인해 발생하는 이상 상태를 탐지할 수 없다는 한계가 있다.

[0008] 이를 위해 최근 머신러닝 기술을 활용하여 VNF의 이상 상태를 탐지하는 연구가 진행되고 있다. 이러한 연구들은 대부분 지도학습(supervised learning), 비지도학습(unsupervised learning), 강화학습(reinforcement learning)과 같은 머신러닝의 3가지 범주 중 지도학습 기반의 알고리즘(Random Forest, Support Vector Machine, Neural Network 등)을 활용하여 이상 상태를 탐지한다. 하지만 대부분의 머신러닝 기반 연구들은 이상 상태를 CPU 및 메모리 사용량와 같은 단순한 측정치를 기준으로 정의하고 있기 때문에 실제 운용되는 서비스 측면에서 SLA(Service Level Agreement) 위반 여부 및 자원 사용 상태를 함께 고려하여 이상 상태를 정의하는 것이 필요하다.

[0009] 또한, 기존의 통계 기반 및 머신러닝 기반의 이상 상태 탐지 방법은 CPU, 메모리, 디스크 어세스(disk access)와 같은 측정치의 임계값을 기준으로 이상 상태를 정의하고 있다. 그리고, 머신러닝 기반의 이상 상태 탐지 방법은 이상 상태를 데이터들의 상호 관계를 통해 학습할 수 있다는 것이다. 하지만 이러한 이상 상태에 대한 정의는 짧은 시간 동안 자원 사용에 대한 측정치가 일시적으로 상승하는 경우, 오탐지를 유발하고 VNF들을 통해 제공되는 서비스에 대한 측면을 고려하지 않는다는 한계점을 지닌다.

발명의 내용

해결하려는 과제

[0011] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은 NFV 환경을 관리하기 위한 VNF의 이상 상태 탐지에 있어, SLA 위반(violation)과 같은 서비스 측면을 함께 고려하여 이상 상태를 정의하여 보다 정확한 이상 탐지 방법을 제공하는 것이다.

[0012] 이를 위해 가상 네트워크에서 자원 사용 및 네트워크 상태, SLA 위반 정보를 모니터링하여 수집한 데이터를 머신러닝에 적용한다. 수집된 데이터는 지도학습 기반의 머신러닝 알고리즘 학습에 사용될 수 있도록 수집된 데이터로부터 의미있는 특성(feature)을 추출하고 데이터를 정상 상태 및 이상 상태로 구분하는 레이블링(labeling) 과정을 거친다.

[0013] 제안하는 방법은 보다 정확한 분류(classification) 정확도와 빠른 훈련을 위해 트리 기반의 알고리즘 중 가장 성능이 우수한 것으로 알려진 XGBoost(eXtreem Gradient Boosting)를 사용한다. 이를 통해 이상 탐지 모델을 생성한 후 모델의 분류 정확도를 검증하고, 이를 이상 탐지 시스템에 활용한다.

[0014] 궁극적으로는 오차가 거의 없는 높은 분류 정확도를 달성함으로써 현재 기존 방법들이 갖는 한계점을 극복하는 이상 탐지 시스템을 구현하는 것에 목표를 두고 있다.

과제의 해결 수단

[0016] 상기 목적을 달성하기 위한 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템은, 물리 네트워크에서 가상화를 통해 구성된 NFV 환경(Network Function Virtualization Infrastructure)의 가상 네트워크에서 동작하는 VNF(Virtualized Network Function)의 이상 상태를 탐지하기 위한 이상 상태 탐지 장치에 있어서, 서비스가 정상적으로 제공되어 생성되는 정상 상태 데이터와 결함 주입 방법을 통해 생성되는 이상 상태 데이터를 모니터링 에이전트와 모니터링 모듈을 통해 실시간으로 수집하고, 수집된 데이터는 시계열(time-series) 데이터 베이스에 저장되고, 시계열 데이터 베이스에 저장된 모니터링 데이터가 이상 상태 여부를 판단하기 위해 데이터 분석부로 전송하는 데이터 수집부; 및 데이터 수집부에서 제공받은 모니터링 데이터를 전처리를 통해 이상 상태 탐지에 필요한 특성을 추출하고, 추출된 특성 데이터를 이상 상태 탐지 모델로 보내면, 이상 상태 탐지 모델은 실시간으로 들어오는 데이터를 분석하여 이상 상태 여부를 판단하고, 이상 상태가 발생한 경우 네트워크 관리자에게 통지하는 데이터 분석부; 를 포함할 수 있다.

[0017] 본 발명의 다른 목적을 달성하기 위한 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법은, 이상 상태 탐지 모델을 학습시키기 위해 NFV 환경(Network Function Virtualization Infrastructure)을 모니터링하

는 NFVI 모니터링 단계, VNF(Virtualized Network Function)의 비정상적인 상태를 발생시키는 결함 주입(fault injection) 단계, 이전 단계에서 수집된 모니터링 데이터를 이상 상태 탐지 모델을 학습시키기에 적합한 형태로 변환하는 전처리(preprocessing) 단계, 및 이상 상태 탐지 알고리즘을 통해 이상 상태 탐지 모델을 학습시키고, 학습된 이상 상태 탐지 모델을 검증한 결과를 비교하여 최적 이상 상태 탐지 모델을 도출하는 이상 상태 탐지 모델 학습 성능 평가 단계; 를 포함할 수 있다.

[0018] 상기 방법은, 이상 상태 탐지 모델 학습 성능 평가 단계에서 도출된 최적 이상 상태 탐지 모델을 기반으로 이상 상태 탐지 알고리즘을 통해 다시 이상 상태 탐지 모델을 학습시키는 피드백 단계를 더 포함할 수 있다.

[0019] NFVI 모니터링 단계는, 모니터링 에이전트(agent)가 가상 네트워크에서 동작하는 각 가상머신의 자원 사용 상태인 모니터링 측정치를 주기적으로 수집하고, 모니터링 모듈(module)이 모니터링 에이전트로부터 수집된 모니터링 측정치 데이터를 수신하고, 수집된 모니터링 측정치 데이터를 시계열 데이터 데이터 베이스에 저장하고, 대쉬보드(dashboard)가 학습을 위한 데이터셋(dataset) 형태로 변환되어 데이터 베이스에 저장된 데이터가 전처리 과정을 거치고 난 후 사용자가 원하는 시각화 형태로 제공받는 단계일 수 있다.

[0020] 결함 주입 단계는, 결함 주입(fault injection)은 실제 운영 환경에서 발생하는 이상 상태의 발생 빈도를 제어하기 위해 사용하는 기술을 이용하여, VNF가 동작하는 가상 네트워크에서 발생 가능한 소프트웨어 및 하드웨어의 이상 상태를 결함 주입 기술을 통해 발생시키는 단계일 수 있다.

[0021] 결함 주입 단계는, VNF가 동작하는 VM에 이상 상태를 발생시키거나, 대량의 트래픽을 전송하여 정상 서비스를 보장할 수 없을 정도의 과부하를 유발하는 결함 주입 기술을 통해 이상 상태를 발생시키는 단계일 수 있다.

[0022] 결함 주입 단계는, VNF가 동작하는 VM에 CPU 부하 및 메모리 부족, 디스크 I/O 엑세스 실패, 네트워크 지연, 네트워크 패킷 손실의 직접적으로 결함을 주입하는 단계이거나, 트래픽 또는 서비스에 대한 접근(access) 및 요청(request)의 허용 범위를 초과하여 들어오는 상황을 발생시켜 패킷 처리의 지연(packet processing delay) 및 커널에 의한 패킷 드롭(packet drop)을 발생하는 단계일 수 있다.

[0023] 전처리 단계는, 모니터링을 통해 수집된 측정값들 중 정상 및 이상 상태를 판별하는데 기준이 되는 값들을 구별하여 선정하고, 수집되는 각 측정치 중 서로 중복되거나 비슷한 특성을 지니는 항목을 제거하여, VNF의 정상 및 이상 상태를 판별하는 특성들을 추출하여 그 데이터를 모델 학습에 사용하는 특성 선택(feature selection) 단계를 포함할 수 있다.

[0024] 전처리 단계는, 추출된 특성 데이터(feature data)를 지도학습 기반의 머신러닝 알고리즘에 사용할 수 있도록 각 시점의 데이터를 정상 상태 및 이상 상태로 분류하는 데이터 레이블링(data labeling) 단계를 포함할 수 있다.

[0025] 전처리 단계는, 결함 주입으로 발생시킨 시스템 및 트래픽의 과부하로 인해 VNF 내부에서 발생하는 SLA 위반을 판단할 수 있는 정보와 서비스의 요청 상태를 기준으로 이상 상태를 정의하고, SLA 위반 및 서비스 요청 실패가 발생하는 경우를 이상 상태로, 이상 상태 이외의 상태를 정상 상태로 레이블링하여 데이터셋을 생성하는 단계일 수 있다.

[0026] 이상 탐지 모델 학습 성능 평가 단계는, 전처리 단계에서 생성된 레이블링 데이터셋을 통해 지도학습 기반의 XGBoost 알고리즘을 사용하여 이상 탐지 모델을 학습시키는 단계일 수 있다.

[0027] 이상 탐지 모델 학습 성능 평가 단계는, 결함 주입 단계 및 전처리 단계에서 SLA 위반 정보 및 응용 서비스 제공 상태를 바탕으로 레이블링된 데이터셋을 통해 XGBoost 알고리즘 기반 학습으로 이상 탐지 모델을 생성하고, 생성된 이상 탐지 모델의 분류 정확도를 검증하고 모델 성능을 평가하는 단계를 포함할 수 있다.

[0028] 모델 학습 단계는, 이상 상태 탐지 학습을 위해 선택된 특성 목록으로 측정 시각, VNF 인스턴스명, CPU - 유휴 시간, CPU - 인터럽트 처리에 소모한 시간, CPU - nice value의 프로세스를 실행하며 소모한 시간, CPU - softirq 처리에 소모한 시간, CPU - hypervisor에 의한 CPU 대기 시간, CPU - kernel 모드에서 소모한 시간, CPU - user 모드에서 소모한 시간, CPU - I/O 대기 시간, 네트워크 인터페이스의 수신 트래픽 대역폭, 네트워크 인터페이스의 송신 트래픽 대역폭, 네트워크 인터페이스의 수신 패킷 수, 네트워크 인터페이스의 송신 패킷 수, Disk - 여유 공간, Disk - 예약된 공간, Disk - 사용 중인 공간, Disk - I/O 읽기, Disk - I/O 쓰기, Disk - I/O 수행 시간, Memory - 여유 공간, Memory - 버퍼된 공간, Memory - 캐시된 공간, Memory - 사용중인 공간, 네트워크 패킷 지연 시간을 포함할 수 있다.

[0029] 모델 학습 단계는, VNF 이상 탐지 모델이 사용하는 XGBoost 알고리즘의 하이퍼 파라미터 값으로 트리 개수, 트

리의 최대 depth, leaf의 최소 observation 수, column 샘플링 비율, 트리당 column 샘플링 비율, early stopping에 사용할 메트릭, early stopping에 사용되는 값, L2 regularization, L1 regularization를 포함할 수 있다.

발명의 효과

[0031] 본 발명은 이러한 한계점을 극복하기 위해 서비스 요청 및 SLA 위반 여부에 따른 이상 상태를 정의하여 문제를 해결하므로, 기존 연구들은 80~90% 사이의 분류 정확도를 보이지만 본 발명에서 이용하는 XGBoost 알고리즘 모델은 기존과 유사한 이상 상태 정의 방법에서도 95% 이상의 높은 분류 정확도를 보이기 때문에 오탐지를 막는데 보다 적합하다. 이는 임계값을 기준으로 이상 상태를 정의하는 방법보다 더 복잡한 SLA 위반 및 서비스 요청 실패 등 서비스 측면에 대한 이상 상태 정의를 했을 경우, 실제 검증이 필요하다는 점을 감안하더라도 기존의 방법보다 높거나 유사한 분류 정확도를 보인다.

[0032] 또한, 본 발명에서는 자원 사용은 물론 SLA 위반과 관련된 다양한 결함 주입 방법을 사용하여 이상 상태를 발생시킴으로써 실제 상황에서 발생 가능한 이상 상태의 다양한 원인을 포함한다.

[0033] 결과적으로, 본 발명을 통해 서비스 측면을 고려하여 이상 상태를 탐지하고 기존보다 높은 분류 정확도를 제공함으로써 보다 정밀한 VNF 이상 상태 탐지 시스템을 구축할 수 있다.

도면의 간단한 설명

[0035] 도 1은 본 발명의 머신러닝 기반 VNF 이상 상태 탐지 시스템의 예시를 나타내는 구성도이다.

도 2는 본 발명의 이상 상태 탐지 모델이 사용하는 XGBoost의 근사 알고리즘 흐름도이다.

도 3 및 도 4는 본 발명의 머신러닝 기반 이상 상태 탐지 방법의 학습 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0036] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0037] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0038] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0039] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0040] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0042] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0044] 도 1은 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템(100)의 예시를 나타내는 구성도이다.

[0045] 도 1을 참조하면, 본 발명에서 제시하는 물리 네트워크(10)에서 가상화를 통해 구성된 NFVI 환경의 가상 네트워크(50)에 적용되는 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템(100)이 개시되어 있다.

[0046] 물리 네트워크(10)에서 가상화를 통해 구성된 NFVI 환경의 가상 네트워크(50)에서 동작하는 본 발명의 VNF의 이상 상태를 탐지하기 위한 이상 상태 탐지 시스템(100)은 데이터 수집부(110)와 데이터 분석부(150)로 구성된다.

[0047] 데이터 수집부(110)는 이상 상태 탐지 모델 학습을 위해 가상 네트워크(50)에서 데이터를 수집하는 부분으로, 서비스가 정상적으로 제공되는 상태의 데이터와 결함 주입 방법을 통해 발생하는 자원 부족 및 네트워크 이상, SLA 위반과 같은 이상 상태의 데이터를 모니터링 에이전트인 컬렉트(collectd)와 모니터링 모듈(111)을 통해 실시간으로 수집한다. 수집된 데이터는 시계열(time-series) 데이터 베이스(113)에 저장되고, 이상 상태를 판단하기 위해 데이터 분석부(150)로 전송된다.

[0048] 데이터 수집부(110)는 모니터링 에이전트(agent) 및 대쉬보드(dashboard)를 더 포함할 수 있다.

[0049] 모니터링 에이전트(agent)가 수집한 모니터링 측정치는 모니터링 모듈(module)(111)을 통해 데이터 베이스(113)에 저장되고 대쉬보드(dashboard)로 시각화하여 구성된다.

[0050] 모니터링 에이전트는 가상 네트워크에서 동작하는 각 가상머신의 자원 사용 상태를 주기적으로 수집한다. 모니터링 에이전트로부터 수집되는 모니터링 측정치는 CPU utilization, memory usage, network traffic load 등 세부항목을 포함하여 모두 73개 항목으로 이루어진다. 모니터링 에이전트는 수집된 측정치인 시계열 모니터링 데이터를 모니터링 모듈(111)로 보낸다.

[0051] 모니터링 모듈(111)은 수집된 시계열 모니터링 데이터를 데이터 베이스(113)에 저장한다.

[0052] 데이터 베이스(113)는 모니터링 모듈(111)에서 수집한 시계열 모니터링 데이터를 저장한다.

[0053] 대쉬보드는 데이터 베이스(113)에 저장된 시계열 모니터링 데이터를 그래프, 표 등과 같이 사용자가 원하는 시각화 형태로 제공한다.

[0054] 데이터 분석부(150)는 데이터 수집부(110)에서 제공받은 모니터링 데이터를 데이터 전처리(151)를 통해 표 1과 같이 이상 상태 탐지에 필요한 특성을 추출하고, 추출된 특성 데이터를 이상 상태 탐지 모델(153)로 보낸다.

[0055] 데이터 전처리(151)는 데이터 베이스(113)에 저장된 모니터링 데이터를 데이터 전처리 과정을 거쳐 학습을 위한 데이터셋(dataset) 형태로 변환된다.

[0056] 이상 상태 탐지 모델(153)은 실시간으로 들어오는 데이터를 분석함으로써 이상 상태 여부를 판단하고, 이상 상태가 발생한 경우 네트워크 관리자(5)에게 통지한다.

[0058] 표 1은 이상 상태 탐지 학습을 위해 선택된 특성 목록이다.

표 1

[0060]

특성 (feature)	설명
Time	측정 시각
instance	VNF 인스턴스명
cpu_idle	CPU - 유휴 시간
cpu_interrupt	CPU - 인터럽트 처리에 소모한 시간
cpu_nice	CPU - nice value의 프로세스를 실행하며 소모한 시간
cpu_softirq	CPU - softirq 처리에 소모한 시간
cpu_steal	CPU - hypervisor에 의한 CPU 대기 시간
cpu_system	CPU - kernel 모드에서 소모한 시간
cpu_user	CPU - user 모드에서 소모한 시간
cpu_wait	CPU - I/O 대기 시간
network_rx_bytes	네트워크 인터페이스의 수신 트래픽 대역폭
network_tx_bytes	네트워크 인터페이스의 송신 트래픽 대역폭
network_rx_packets	네트워크 인터페이스의 수신 패킷 수
network_tx_packets	네트워크 인터페이스의 송신 패킷 수
disk_free	Disk - 여유 공간

disk_reserved	Disk - 예약된 공간
disk_used	Disk - 사용 중인 공간
disk_read	Disk - I/O 읽기
disk_write	Disk - I/O 쓰기
disk_Io_time	Disk - I/O 수행 시간
mem_free	Memory - 여유 공간
mem_buffered	Memory - 버퍼된 공간
mem_cashed	Memory - 캐시된 공간
mem_used	Memory - 사용중인 공간
hop-by-hop latency	네트워크 패킷 지연 시간

[0062] 본 발명에서 제안하는 방법을 통해 VNF 이상 탐지 모델(153)을 학습시키기 위해 사용하는 데이터셋의 정상 및 이상 데이터 레이블링은 다음과 같이 이루어진다. 먼저 데이터셋은 앞서 설명한 바와 같이 수집된 모니터링 데이터를 모델 학습에 적합한 형태로 변환하여 생성되며, 이를 위해 모니터링 과정에서 수집된 각 메트릭 중 이상 상태를 구별하기 위한 기준과 가장 관련이 있는 메트릭을 선별한다. 이 과정은 각 메트릭의 상호 관계(correlation)를 고려하여 이루어진다. 다음으로 데이터의 정상 및 이상 상태 레이블링의 경우, CPU 사용량와 같은 메트릭을 레이블링 기준으로 정한다면 많은 오탐을 유발한다. 따라서 본 발명에서는 VNF의 성능 문제(performance bottleneck)가 발생하거나 SLA 위반이 발생했을 경우를 이상 상태로 정의한다.

[0063] VNF의 성능 문제는 주로 VNF의 과부하 혹은 결함 주입으로 인해 사용 가능한 시스템 리소스가 부족하게 되어 VNF 내부의 패킷 손실(packet loss)을 유발하기 때문에 본 발명에서는 패킷 손실율이 1% 이상일 때를 이상 상태로 정의하여 어떤 VNF에 이상이 발생했는지(root cause localization)를 탐지한다. SLA 위반의 경우 제공되는 서비스마다 그 기준이 다르지만 일반적으로 평균 서비스 시간(average response time) 및 서비스 요청에 대한 실패율(request failure rate)을 포함하기 때문에 이러한 지표를 기준으로 이상 상태를 정의하며, 이와 더불어 각 서비스에 부합하는 SLA 위반 기준을 이상 상태로 정의한다. 예를 들어, 웹 호스팅 서비스의 경우 SLA 위반은 평균 응답 시간이 0.5초, 1초, 혹은 2초 이상이 소요되는 경우, 그리고 서비스 요청에 대한 실패율이 0.1%, 1%, 2% 이상일 때를 SLA 위반으로 정의하고 있다(GFD-R. 192-Web Service Agreement Specification기준).

[0064] 또한, 본 발명에서 사용하는 XGBoost 알고리즘은 다수의 모델을 학습시키고 결합함으로써 단일 모델을 통해 학습시켰을 때보다 우수한 성능을 가지는 모델을 얻는 앙상블 학습 기법을 기반으로 한다. XGBoost는 앙상블 학습 기법 중 부스팅(boosting) 기법에 해당하는 알고리즘으로, 부스팅(boosting) 기법은 이전에 학습한 모델에서 분류 오류가 있는 데이터에 대하여 가중치를 높여 다음 모델 학습에서 분류 정확도를 높인다. 부스팅(boosting) 기법 기반의 알고리즘 중 일반적으로 널리 사용되는 GBM과는 달리 XGBoost는 장점을 지닌다.

[0066] 도 2는 본 발명의 이상 상태 탐지 모델이 사용하는 XGBoost의 근사 알고리즘 흐름도이다.

[0067] 도 2를 참조하면, 본 발명의 이상 탐지 모델이 사용하는 XGBoost의 알고리즘은 다음의 수학식 1 내지 수학식 4로 설명된다.

[0068] 먼저 XGBoost는 GBM이 가지는 과적합 문제를 해결하기 위해 수학식 1과 같이 정규화를 적용한 목적 함수(objective function)를 통해 과적합을 방지한다.

수학식 1

[0070] $$L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{n} \Omega(f_i)$$

[0071] l: 손실 함수 ($\hat{y}i^t$: 예측값, yi: 실제 결과값)

[0073] 수학식 1에서 첫 항 (l)은 손실 함수(differentiable convex loss function)로, 이는 i번째 인스턴스의 예측값 $\hat{y}_i$와 실제 결과값 y_i 의 차이를 나타낸다. 두 번째 항 (Ω)은 각 트리의 복잡도 나타내는 정규화 기법으로 각 트리에 대해 수학식 2와 같이 트리의 리프(leaf) 개수 T와 리프의 가중치 벡터의 노름(norm) $\|w\|^2$을 손실 함수에 더해줌으로써, 목적 함수의 최소화 과정에서 모델의 복잡도를 제어하여 과적합 문제를 해결한다.

수학식 2

[0075] $$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$$

[0076] γT: 트리의 리프 개수

[0077] $\|w\|^2$: 리프의 가중치 벡터의 노름 (norm)

[0079] 전술한 목적 함수와 더불어 XGBoost는 과적합 문제 해결을 위해 Shrinkage 스케일링(scaling)과 컬럼 서브샘플링(column subsampling)을 사용한다. Shrinkage 스케일링은 부스팅(boosting) 기반 트리의 각 단계에서 새롭게 추가되는 가중치에 대한 스케일링을 적용하여 확률적인(stochastic) 최적화 과정에서 새로운 트리에 대한 기존의 트리나 리프의 영향을 감소시킨다. 서브샘플링(column subsampling)은 기존 열(row) 기반 서브샘플링(subsampling)보다 과적합을 방지하며 학습 속도를 향상시킨다.

[0080] 또한 기존 GBM은 각 특성마다 모든 분할점에 대한 최적화 지점을 탐색하는 과정에서 탐욕 알고리즘(greedy algorithm)을 사용하기 때문에 높은 분류 정확도를 제공하지만 학습 시간이 오래 걸린다는 제약이 존재한다. 이에 반해 XGBoost는 최적화된 분할점 탐색을 위해 도 2와 같은 근사 알고리즘을 사용한다. 근사 알고리즘(approximate algorithm)은 각 특성에 대한 후보 분할점을 설정하고(S30), 특성 분포의 분위수(quantile)에 따라 분할된 구간별 손실 함수의 기울기 벡터를 합산한다(S40). 이를 기반으로 분할 최적화에 대한 점수를 계산하고 분할점 설정을 최종적으로 확정할지 여부를 결정한다(S50).

[0081] 각 특성에 대한 후보 분할점을 적절하게 설정하기 위해 XGBoost의 근사 알고리즘은 가중치를 적용한 분위수 스케치 방법(weighted quantile sketch)(S10)과 희소성 인식 방법(sparsity-aware split finding)(S20)을 적용하여 후보 분할점을 탐색한다. 분위수 스케치 방법(S10)은 수학식 3과 같이 특성 k에 대한 데이터를 $1/\epsilon$ 로 분할하는 근사 계수 ϵ 를 통해 데이터를 균일하게 분등하는 분할점 $\{s_{k,1}, s_{k,2}, \cdots, s_{k,l}\}$을 찾는다.

수학식 3

[0083] $$|r_k(s_{k,j}) - r_k(s_{k,j+1})| < \epsilon$$

[0084] ϵ: 근사 계수 (approximation factor)

[0085] $s_{k,j}$: 특성 k에 대한 j번째 분할점

[0087] 데이터를 균일하게 분할하기 위해 각 분할점보다 작은 데이터의 비율을 나타내는 함수 r_k 는 수학식 4와 같이 정의하여 데이터의 분할에 사용된다. 이 때, D_k는 특성 k에 대하여 가중치를 적용한 데이터셋을, h는 데이터의 가중치를 의미한다. XGBoost는 상기 분위수 스케치 방법을 통해 가중치가 있는 데이터에 대해 정확도를 유지하며 분할점을 찾는다.

수학식 4

[0089] $$r_k(z) = \frac{1}{\sum_{(x,h)\in D_k} h}\sum_{(x,h)\in D_k, x<z} h$$

[0090] D_k: 특성 k 에 대한 데이터셋

[0091] h: 데이터에 대한 가중치

[0093] 희소성 인식 방법(S20)은 데이터 수집 과정에서 값이 누락되어 결측치가 발생하거나 데이터가 희소한(sparse) 경우 결측 데이터 및 희소성 데이터를 고려하여 분할점을 찾는다. 예를 들어 각 트리의 노드에 기본 분류 방향을 설정하여 데이터에 값이 누락된 경우, 누락된 값을 기본 분류 방향으로 분류한다.

[0095] 표 2는 제안하는 VNF 이상 탐지 모델이 사용하는 XGBoost 알고리즘의 하이퍼 파라미터 값이다.

표 2

[0097]

하이퍼 파라미터	값	설명
ntrees	111	트리 개수
max_depth	5	트리의 최대 depth
min_rows	3	leaf의 최소 observation 수
col_sample_rate	0.8	column 샘플링 비율
col_sample_rate_per_tree	0.8	트리당 column 샘플링 비율
stopping_metric	Logloss	early stopping에 사용할 메트릭
stopping_tolerance	0.0045469579205	early stopping에 사용되는 값
reg_lambda	0.001	L2 regularization
reg_alpha	1	L1 regularization

[0099] NFV 환경에서 결함 주입 방법을 통해 생성한 데이터셋과 XGBoost 알고리즘을 기반으로 이상 탐지 모델을 학습시키기 위해 본 발명에서는 표 2와 같은 하이퍼 파라미터를 이용하여 이상 탐지 모델의 성능을 최적화한다.

[0100] 이를 기반으로 생성된 이상 상태 탐지 모델의 성능 검증을 위해 데이터를 레이블링하고(S400), 레이블링된 데이터를 75%, 25%의 학습 데이터셋(training dataset)와 테스트 데이터셋(test dataset)으로 나누고 이상 상태 탐지 모델을 학습하여, 학습 데이터셋을 통해 학습된 이상 상태 탐지 모델의 성능을 5겹 교차검증(5-fold cross validation) 방법으로 평가한다. 이상 상태 탐지 모델의 평가를 위한 항목으로는 정확도(accuracy), 정밀도(precision), 재현율(recall), F-Measure(F1 score) 등을 사용한다. 그 후, 이상 상태 탐지 모델 학습에 관여하지 않은 테스트 데이터셋을 통해 최종적으로 이상 상태 탐지 모델의 성능을 평가한다.

[0102] 도 3 및 도 4는 본 발명의 머신러닝 기반 이상 상태 탐지 방법의 학습 흐름도이다.

[0103] 도 3 및 도 4를 참조하면, 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법은, 이상 상태 탐지 모델을 학습시키기 위해 NFV 환경(Network Function Virtualization Infrastructure)을 모니터링하는 NFVI 모니터링 단계(S100), VNF(Virtualized Network Function)의 비정상적인 상태를 발생시키는 결함 주입(fault injection) 단계(S200), 이전 단계에서 수집된 모니터링 데이터를 이상 상태 탐지 모델을 학습시키기에 적합한 형태로 변환하는 전처리(preprocessing) 단계(S300), 및 이상 상태 탐지 알고리즘을 통해 이상 상태 탐지 모델을 학습시키고, 학습된 이상 상태 탐지 모델을 검증한 결과를 비교하여 최적 이상 상태 탐지 모델을 도출하는 이상 상태 탐지 모델 학습 성능 평가 단계(S400)를 포함한다.

[0104] 여기서, 전처리 단계(S300) 단계는, 특성(feature) 선택 단계(S310), 데이터 레이블링 단계(S350)를 포함하고, 이상 상태 탐지 모델 학습 성능 평가 단계(S400) 단계는 모델 학습 단계(S410), 모델 성능 평가 단계(S450)를 포함한다.

[0105] 여기서, 이상 상태 탐지 모델 학습 성능 평가 단계(S400)는, 모델 성능 평가 단계(S450)에서 도출된 최적 이상 상태 탐지 모델을 기반으로 이상 상태 탐지 알고리즘을 통해 다시 이상 상태 탐지 모델을 학습시키는 단계(S410)가 재반복되는 피드백 단계(S470)를 더 포함한다.

[0107] 전술한 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템을 이용한 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 방법을 설명하면, 발명의 이상 탐지 모델 생성 방법은 크게 4가지 단계로 구성된다. 첫 번째 단계는 NFVI(NFV Infrastructure) 모니터링 단계(S100)로, 이상 상태 탐지 모델을 학습시키기 위해 NFVI 환경을 모니터링하고, 두 번째 단계인 결함 주입(fault injection) 단계(S200)에서는 VNF의 비정상적인 상태를 발생시키며, 세 번째 단계인 전처리(preprocessing) 단계(S300)에서는 이전 단계에서 수집된 모니터링 데이터를 머신러닝 모델을 학습시키기에 적합한 형태로 변환하기 위해 특성 선택 단계(S310)와 데이터 레이블링 단계(S350)를 진행하고, 마지막으로, 이상 탐지 모델 학습 성능 평가 단계(S400)에서는 XGBoost 알고리즘을 통해 이상 상태 탐지 모델을 학습(S410)시키고, 각 모델을 검증한 결과를 비교하여 최적의 모델을 도출하는 모델 성능 평가(S450) 단계를 진행한다.

[0108] NFVI 모니터링 단계(S100)는 일반적으로 모니터링 에이전트(agent)가 수집한 모니터링 측정치는 모니터링 모듈(module)(111)을 통해 데이터 베이스(113)에 저장되고 대쉬보드(dashboard)로 시각화하여 구성된다. 모니터링 에이전트는 가상 네트워크에서 동작하는 각 가상머신의 자원 사용 상태를 주기적으로 수집한다. 모니터링 에이전트로부터 수집되는 모니터링 측정치는 CPU utilization, memory usage, network traffic load 등 세부항목을 포함하여 모두 73개 항목으로 이루어진다. 모니터링 에이전트는 데이터를 모니터링 모듈(111)로 보내고, 모니터링 모듈(111)은 수집된 데이터를 시계열 데이터 데이터 베이스(113)에 저장한다. 저장된 데이터는 전처리 과정을 거친 후, 학습을 위한 데이터셋(dataset) 형태로 변환된다. 대쉬보드를 통해 데이터 베이스(113)에 저장된 데이터를 그래프, 표 등과 같이 사용자가 원하는 시각화 형태로 제공받는다.

[0109] 결함 주입(fault injection) 단계(S200)는 실제 운영 환경에서 매우 드물게 일어나는 이상 상태의 발생 빈도를 제어하기 위해 사용하는 기술이다. VNF가 동작하는 가상 네트워크에서 발생 가능한 다양한 소프트웨어 및 하드웨어의 이상 상태를 결함 주입 기술을 통해 발생시킨다. 결함 주입 기술을 통해 이상 상태를 발생시키는 데에는 크게 두 가지 방법이 가능하다. 첫째는 VNF가 동작하는 VM에 이상 상태를 발생시키는 것이고, 둘째는 대량의 트래픽을 전송함으로써 올바른 서비스를 보장할 수 없을 정도의 과부하를 유발하는 것이다. 첫 번째 방법은 VNF가 동작하는 VM에 직접적으로 결함을 주입한다. 이는 CPU 부하 및 메모리 부족, 디스크 I/O 엑세스 실패, 네트워크 지연, 네트워크 패킷 손실 등을 발생시킨다. 두 번째 방법은 대량의 트래픽을 통해 네트워크 과부하를 발생시켜 VNF가 들어오는 패킷을 처리하는데 많은 시스템 자원 및 시간을 소요하게 한다. 예를 들어, 트래픽 또는 서비스에 대한 접근(access) 및 요청(request)이 과다하게 들어오는 상황을 발생시켜 패킷 처리의 지연(packet processing delay) 및 커널에 의한 패킷 드롭(packet drop)을 발생시킨다.

[0110] 전처리 단계(S300)는 특성 선택(feature selection) 단계(S310)와 데이터 레이블링(data labeling) 단계(S350)로 구성된다. 먼저, 특성 선택 단계(S310)는 모니터링을 통해 수집된 측정값들 중 정상 및 이상 상태를 판별하는데 기준이 되는 값들을 구별하여 선정하는 단계이다. 이 S310 단계에서는 수집되는 각 측정치 중 서로 중복되거나 비슷한 특성을 지니는 항목을 제거한다. 이 과정을 통해 VNF의 정상 및 이상 상태를 판별하는 특성들을 추출하여 그 데이터를 모델 학습에 사용한다. 데이터 레이블링 단계(S350)는 추출된 특성 데이터(feature data)를 지도학습 기반의 머신러닝 알고리즘에 사용할 수 있도록 각 시점의 데이터를 정상 상태 및 이상 상태로 분류하는 단계이다. 이상 상태는 결함 주입으로 발생시킨 시스템 및 트래픽의 과부하로 인해 VNF 내부에서 발생하는 SLA 위반을 판단할 수 있는 정보와 서비스의 요청 상태를 기준으로 정의한다. 즉, SLA 위반 및 서비스 요청 실패가 발생하는 경우를 이상 상태로, 나머지를 정상 상태로 레이블링하여 데이터셋을 생성한다.

[0111] 마지막으로, 이상 탐지 모델 학습 성능 평가 단계(S400)는 전처리 단계(S300)에서 생성된 레이블링 데이터셋을 통해 지도학습 기반의 XGBoost 알고리즘을 사용하여 이상 탐지 모델을 학습시킨다(S410). XGBoost는 결정 트리(Decision Tree)에 기반한 머신러닝 알고리즘으로, 결정 트리 기반의 알고리즘은 이미지나 텍스트 등의 비정형 데이터의 예측 문제에서 좋은 성능을 보이는 신경망(Neural Network) 기반의 알고리즘과는 달리 정형 데이터의 분류 및 예측에서 보다 우세한 성능을 보인다. 특히, XGBoost는 일반적으로 널리 사용되는 부스팅(boosting) 기법 기반의 알고리즘인 GBM(Gradient Boosting Machine)과 같은 독립적인 트리를 반복적으로 학습시키는 방식을 취하지만, GBM이 가지는 과적합(overfitting) 문제를 해결하고 자원 사용 및 학습 속도 측면에서 GBM 보다 우수한 성능을 보인다. 이상 탐지 모델 학습 성능 평가 단계(S400)에서는, 결함 주입(S200) 및 전처리 단계(S300)에서 SLA 위반 정보 및 응용 서비스 제공 상태를 바탕으로 레이블링된 데이터셋을 통해 XGBoost 알고리즘 기반 학습으로 이상 탐지 모델을 생성하고(S410), 생성된 이상 탐지 모델의 분류 정확도를 검증하고 이상 탐지 모델 성능을 평가하고(S450), 그리고 이상 탐지 모델 성능 평가 단계(S450)의 결과로 생성된 최적 이상 탐지 모델을 다시 이상 상태 탐지 모델 학습 단계(S410)에 피드백(S470)하는 일련의 프로세스로 동작하는 VNF의 이상 탐지 시스템(100)을 구축하여 NFV 환경 관리에 활용한다.

[0113] 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법은 이상 상태를 데이터들의 상호 관계를 통해 학습할 수 있다는 것이다. 하지만 기존 머신러닝 기반의 이상 상태 탐지 방법은 이상 상태에 대한 정의에 있어 CPU 및 메모리 등과 같은 측정치의 임계값을 기준으로 이상 상태를 정의하고 있기 때문에 많은 오탐지를 유발하고 실제 제공되는 서비스의 상태를 고려하지 않는다는 한계점을 지닌다.

[0114] 따라서 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법은 이러한 한계점을 극복하기 위해 서비스 요청 및 SLA 위반 여부에 따른 이상 상태를 정의하여 문제를 해결한다. 기존 연구들은 80~90% 사이의 분류 정확도를 보이지만 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법에서 이용하는 XGBoost 알고리즘 모델은 기존과 유사한 이상 상태 정의 방법에서도 95% 이상의 높은 분류 정확도를 보이기 때문에 오탐지를 막는데 보다 적합하다. 이는 임계값을 기준으로 이상 상태를 정의하

는 방법보다 더 복잡한 SLA 위반 및 서비스 요청 실패 등 서비스 측면에 대한 이상 상태 정의를 했을 경우, 실제 검증이 필요하다는 점을 감안하더라도 기존의 방법보다 높거나 유사한 분류 정확도를 보일 것으로 예상된다.

[0115] 또한 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법에서는 자원 사용은 물론 SLA 위반과 관련된 다양한 결함 주입 방법을 사용하여 이상 상태를 발생시킴으로써 실제 상황에서 발생 가능한 이상 상태의 다양한 원인을 포함한다. 결과적으로 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법은 이상 상태를 탐지하고 기존보다 높은 분류 정확도를 제공하는 서비스 측면을 고려함으로써 보다 정밀한 VNF 이상 상태 탐지 시스템을 구축할 수 있다.

[0117] 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법은 현재 NFV 환경이 고도화되고 복잡해짐에 따라 발생하는 NFV 환경의 관리 문제를 해결하기 위해 머신러닝 기반의 VNF의 이상 상태 탐지 모델을 생성하는 방법을 정의하고, 이를 통해 생성된 모델을 NFV 환경에 적용하여 실제 동작 중인 VNF의 이상 상태를 탐지하는 방법을 제안한다.

[0118] 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법에서 사용하는 이상 탐지 모델 학습 방법은 XGBoost와 같이 기존의 방법에 사용되지 않은 새로운 머신러닝 알고리즘들을 통해 가장 좋은 정확도를 가지는 최적의 모델을 생성할 수 있다.

[0119] 또한, 기존 시스템이 CPU, 메모리와 같은 단순한 측정치를 기준으로 이상 상태를 탐지하는 방법을 개선하여 본 발명의 가상 네트워크 관리를 위한 머신 러닝 기반 VNF 이상 탐지 시스템 및 방법은 SLA 위반 여부를 포함한 서비스의 상태를 고려하여 이상 상태를 정의함으로써 보다 정밀한 이상 탐지 시스템을 실현할 수 있다.

[0121] 본 발명의 실시예에 따른 방법의 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 정보가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.

[0122] 또한, 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

[0123] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.

[0124] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그머블 게이트 어레이는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

[0125] 이상 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

[0127] 5 : 네트워크 관리자

10 : 물리 네트워크

50 : 가상 네트워크

100 : 이상 상태 탐지 시스템

110 : 데이터 수집부

111 : 모니터링 모듈

113 : 데이터 베이스

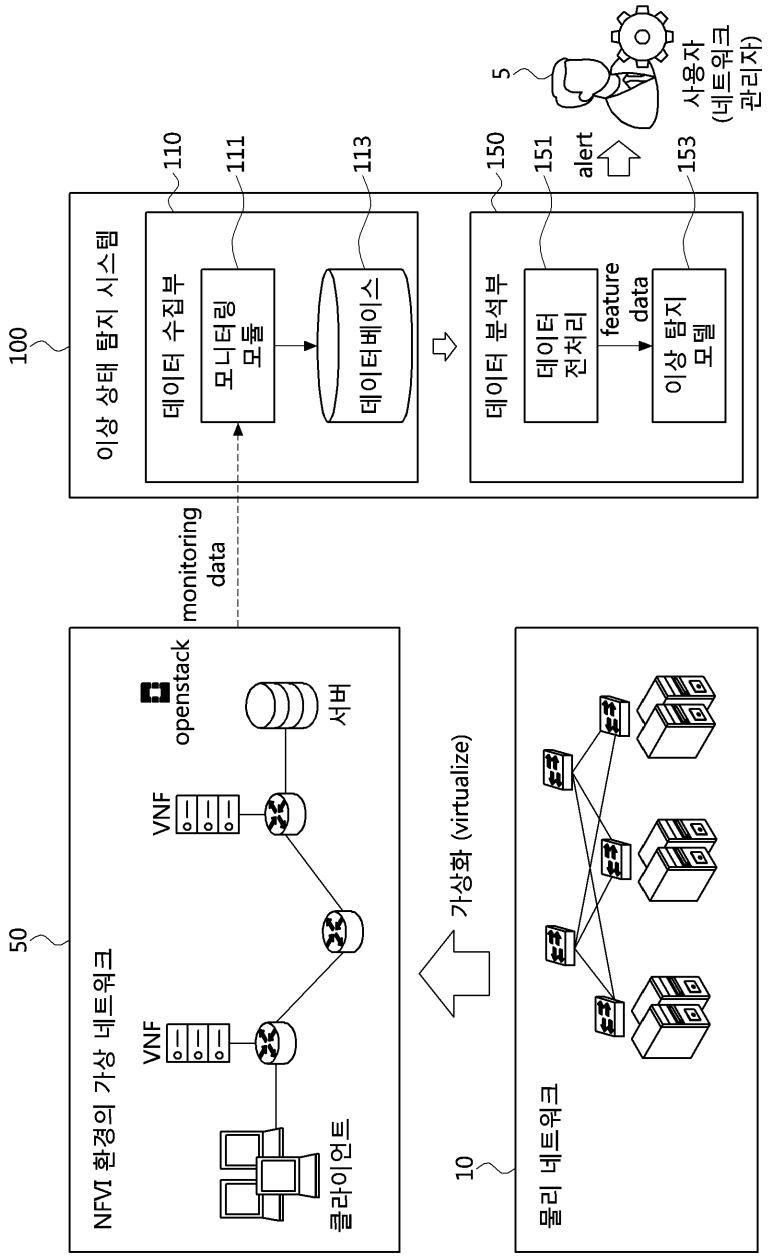
150 : 데이터 분석부

151 : 데이터 전처리

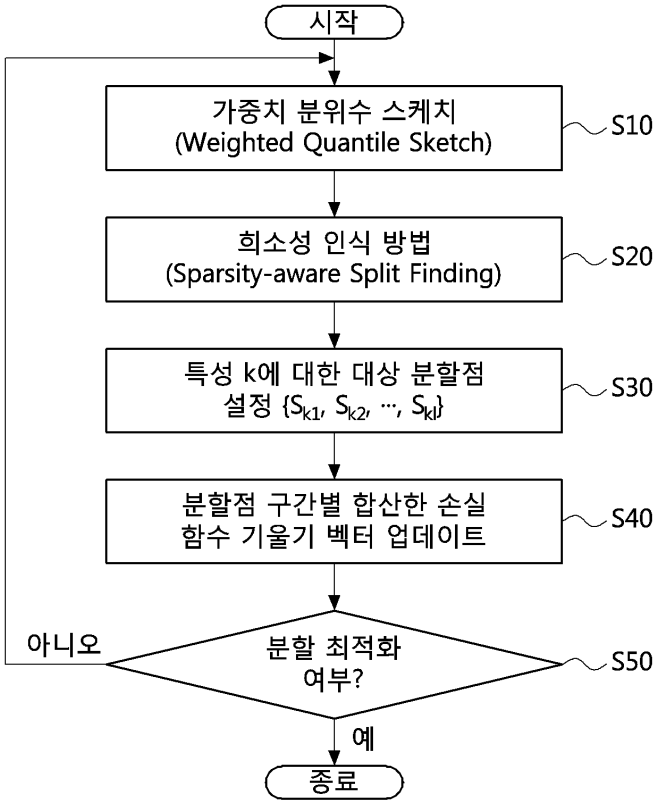
153 : 이상 상태 탐지 모델

도면

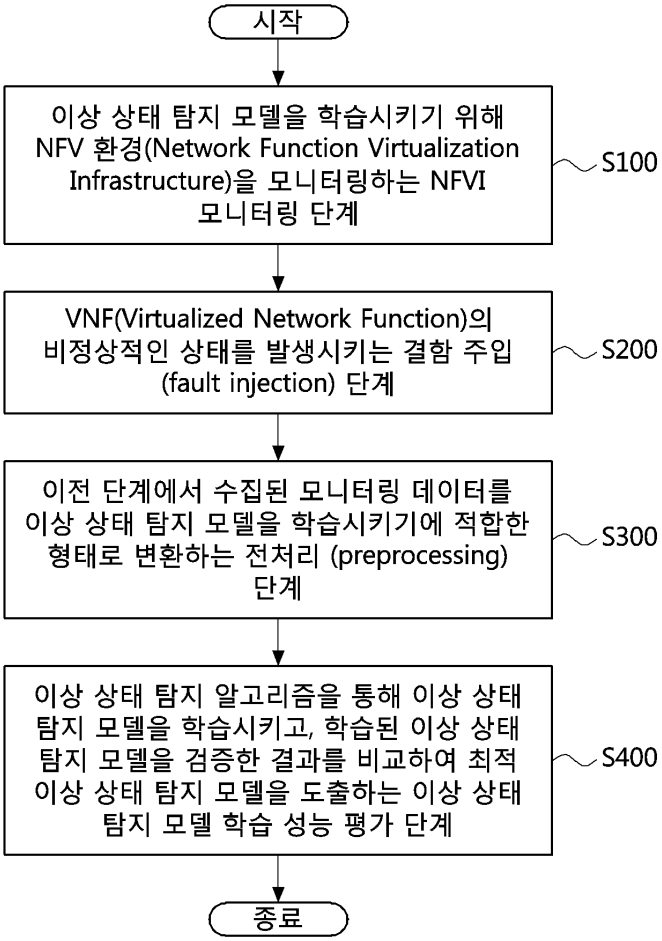
도면1



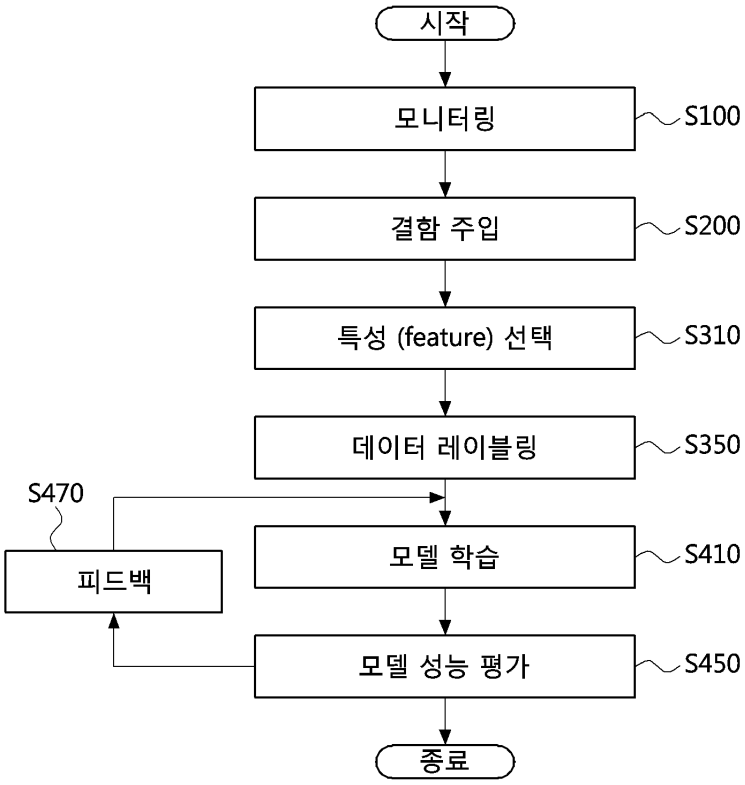
도면2


시작
가중치 분위수 스케치
(Weighted Quantile Sketch)
S10
희소성 인식 방법
(Sparsity-aware Split Finding)
S20
특성 k에 대한 대상 분할점
설정 {Sk1, Sk2, ···, Skl}
S30
분할점 구간별 합산한 손실
함수 기울기 벡터 업데이트
S40
분할 최적화
여부?
S50
아니오
예
종료

도면3


시작
이상 상태 탐지 모델을 학습시키기 위해 NFV 환경(Network Function Virtualization Infrastructure)을 모니터링하는 NFVI 모니터링 단계
S100
VNF(Virtualized Network Function)의 비정상적인 상태를 발생시키는 결함 주입(fault injection) 단계
S200
이전 단계에서 수집된 모니터링 데이터를 이상 상태 탐지 모델을 학습시키기에 적합한 형태로 변환하는 전처리 (preprocessing) 단계
S300
이상 상태 탐지 알고리즘을 통해 이상 상태 탐지 모델을 학습시키고, 학습된 이상 상태 탐지 모델을 검증한 결과를 비교하여 최적 이상 상태 탐지 모델을 도출하는 이상 상태 탐지 모델 학습 성능 평가 단계
S400
종료

도면4


시작
모니터링
S100
결함 주입
S200
특성 (feature) 선택
S310
데이터 레이블링
S350
S470
피드백
모델 학습
S410
모델 성능 평가
S450
종료