



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년03월30일
(11) 등록번호 10-2516187
(24) 등록일자 2023년03월27일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) G06N 5/02 (2023.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06N 5/02 (2023.01)
(21) 출원번호 10-2020-0154336
(22) 출원일자 2020년11월18일
심사청구일자 2020년11월18일
(65) 공개번호 10-2022-0067758
(43) 공개일자 2022년05월25일
(56) 선행기술조사문헌
KR1020180120558 A*
KR1020190070702 A*
KR1020190001501 A
KR1020170125238 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
(주)글루시스
경기도 안양시 동안구 시민대로327번길 11-31 ,
5층(관양동, 파낙스알앤디센터)
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
에프에이리눅스 주식회사
경기도 의왕시 이미로 40, 비동 1001호 1002호 (포일동, 인덕원아이티밸리)
(72) 발명자
노재춘
서울특별시 광진구 능동로 209, 대양AI센터 442호(군자동, 세종대학교)
권세훈
서울특별시 광진구 광나루로16길 9-10
(뒤편에 계속)
(74) 대리인
민영준

전체 청구항 수 : 총 5 항

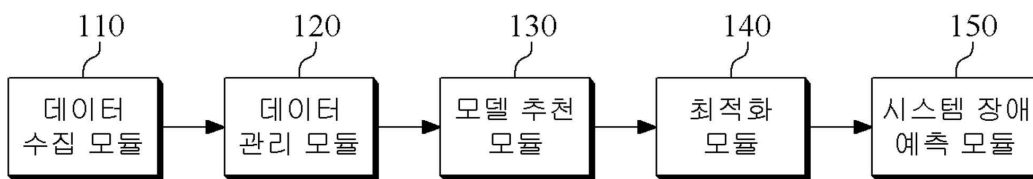
심사관 : 박승철

(54) 발명의 명칭 시스템 장애 예측 방법 및 시스템

(57) 요약

기계 학습 기반으로 시스템의 장애를 예측하는 방법 및 시스템이 개시된다. 개시된 시스템 장애 예측을 위한 학습 방법은 타겟 시스템에 대한 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계; 상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계; 및 제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계를 포함한다.

대표도 - 도1



(72) 발명자

서형준

서울특별시 노원구 동일로191가길 59, 103동 1705호(공릉동, 공릉동 신도1차아파트)

박성순

경기도 군포시 수리산로 244, 995동 901호(산본동, 한양백두아파트)

박문식

경기도 부천시 역곡로85번길 3(역곡동)

김경표

광주광역시 동구 용산3길 17(용산동, 모아엘가 에듀파크)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711117149
과제번호	2020-0-00104-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대 엣지컴퓨팅 시스템 기술개발사업
연구과제명	저지연 입출력 집약적 엣지 데이터 처리를 위한 스토리지 모듈 기술 개발
기 여 율	1/1
과제수행기관명	(주)글루시스
연구기간	2020.04.01 ~ 2020.12.31

명세서

청구범위

청구항 1

장애 예측 시스템에 의해 수행되는 시스템 장애 예측을 위한 학습 방법에 있어서,

타겟 시스템에 대한 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계;

상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계; 및

상기 제1훈련 데이터와 다른 시점에 수집된 제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계를 포함하며,

상기 재학습 장애 예측 모델 후보군을 결정하는 단계는

상기 제1 및 제2훈련 데이터의 패턴 유사도를 판단하는 단계; 및

상기 제1 및 제2훈련 데이터의 패턴 유사도가 임계값 이상인 경우, 상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 상기 복수의 장애 예측 모델에서 상기 재학습 장애 예측 모델 후보군을 결정하는 단계

를 포함하는 시스템 장애 예측을 위한 학습 방법.

청구항 2

삭제

청구항 3

제 1항에 있어서,

상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계는

전처리된 상기 제2훈련 데이터를 이용하여 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계; 및

상기 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 저장하는 단계

를 포함하는 시스템 장애 예측을 위한 학습 방법.

청구항 4

제 1항에 있어서,

상기 시스템 데이터는

미리 설정된 수집 주기에 따라서 수집된, 상기 타겟 시스템의 시계열 형태의 자원 정보

를 포함하는 시스템 장애 예측을 위한 학습 방법.

청구항 5

삭제

청구항 6

삭제

청구항 7

장애 예측 시스템에 의해 수행되는 시스템 장애 예측 방법에 있어서,

타겟 시스템에 대한 훈련용 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계;

상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계;

제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계; 및

상기 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템의 장애 발생을 예측하는 단계를 포함하며,

상기 타겟 시스템에 대한 장애를 예측하는 단계는

상기 타겟 시스템에 대한 타겟 시스템 데이터를 수집하는 단계;

상기 타겟 시스템 데이터 및 상기 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템의 장애 발생 여부를 예측하는 단계; 및

상기 타겟 시스템에 대한 장애가 발생할 것으로 예측된 경우, 상기 베스트 장애 예측 모델의 성능 평가값을 이용하여, 상기 베스트 장애 예측 모델의 장애 발생 예측 정확도를 제공하는 단계를 포함하며,

상기 장애 발생 예측 정확도는

상기 베스트 장애 예측 모델의 성능 평가값을 상기 베스트 장애 예측 모델의 최적화 횟수로 나눈 값에 대응되는 시스템 장애 예측 방법.

청구항 8

장애 예측 시스템에 의해 수행되는 시스템 장애 예측 방법에 있어서,

타겟 시스템에 대한 훈련용 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계;

상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계;

제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계; 및

상기 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템의 장애 발생을 예측하는 단계를 포함하며,

상기 타겟 시스템에 대한 장애를 예측하는 단계는

상기 타겟 시스템에 대한 타겟 시스템 데이터를 수집하는 단계;

상기 타겟 시스템 데이터 및 상기 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템의 장애 발생 여부를 예측하는 단계; 및

상기 타겟 시스템에 대한 장애가 발생할 것으로 예측된 경우, 상기 베스트 장애 예측 모델의 성능 평가값을 이용하여, 상기 베스트 장애 예측 모델의 장애 발생 예측 정확도를 제공하는 단계를 포함하며,

상기 장애 발생 예측 정확도는

장애 예측 이력 정보에 따라서 조절되며,

상기 장애 예측 이력 정보는

상기 베스트 장애 예측 모델의 장애 발생 예측 내용과, 상기 타겟 시스템의 장애 발생 사이의 일치 비율, 상기 베스트 장애 예측 모델의 장애 발생 예측 시점과 상기 타겟 시스템의 장애 발생 시점 사이의 시간 차이값을 포함하며,

상기 장애 발생 예측 정확도는

상기 일치 비율 및 상기 시간 차이값에 비례하는

시스템 장애 예측 방법.

청구항 9

삭제

청구항 10

삭제

발명의 설명

기술 분야

[0001] 본 발명은 시스템 장애 예측 방법 및 시스템에 관한 것으로서, 더욱 상세하게는 기계 학습 기반으로 시스템의 장애를 예측하는 방법 및 시스템에 관한 것이다.

배경 기술

[0003] 다양한 기업에서 은행, 통신, 클라우드 등의 IT 서비스를 지원함으로써 컴퓨터 시스템이 중요해졌으며, 시스템의 가용성이 기업에 수익성에 상당한 영향을 미칠 수 있기 때문에, 장애 예측을 통한 가용성 확보가 중요해졌다. 시스템 자원의 장애로 인한 피해를 최소화하기 위해 가동 중인 시스템의 장애를 예측하는 온라인 장애 예측을 통한 다양한 연구들이 수행되고 있다.

[0004] 온라인 장애 예측은 기계 학습, 통계 분석, 패턴 인식 등 여러 기술들을 통해 컴퓨터 시스템의 장애를 예측하고, 이를 통해 적절한 대처를 할 수 있다. 최근에는 기계 학습을 이용한 방법들이 좋은 성능을 보여주고 있다.

[0005] 기계 학습을 사용하여 장애 예측 모델을 만드는 경우 데이터 가공, 특징 추출, 특징 변환, 특징 선택 및 하이퍼파라미터 튜닝 등 다양한 기법들이 사용된다. 장애 예측 모델 최적화 및 생성 과정에서는 반복적인 기계 학습을 통한 성능 평가와 비교가 이루어지므로 많은 비용이 소요된다. 다양한 장애 예측 모델이 이용되는 경우, 보다 많은 비용과 시간이 소요된다.

[0006] 관련 선행문헌으로 특허 문헌인 대한민국 등록특허 제10-1984730호, 제10-2149930호가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 시스템의 장애 예측에 소요되는 비용과 시간을 줄일 수 있는 시스템 장애 예측 방법 및 시스템을 제공하기 위한 것이다.

과제의 해결 수단

[0010] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 타겟 시스템에 대한 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계; 상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계; 및 제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계를 포함하는 시스템 장애 예측을 위한 학습 방법이 제공된다.

[0011] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 타겟 시스템에 대한 훈련용 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여, 복수의 장애 예측 모델을 학습하는 단계; 상기 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정하는 단계; 제2훈련 데이터를 이용하여, 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하는 단계; 및 상기 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템의 장애 발생을 예측하는 단계를 포함하는 시스템 장애 예측 방법이 제공된다.

[0012] 또한 상기한 목적을 달성하기 위한 본 발명의 또 다른 실시예에 따르면, 타겟 시스템에 대한 시스템 데이터를 수집하는 데이터 수집 모듈; 상기 시스템 데이터를 전처리하는 데이터 관리 모듈; 미리 학습된 복수의 장애 예측 모델에 대한 성능 평가값에 기반하여, 상기 장애 예측 모델 중에서 재학습 장애 예측 모델 후보군을 결정하는 모델 추천 모듈; 상기 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하고, 상기 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 결정하는 최적화 모듈; 및 상기 시스템 데이터 및 상기 베스트 장애 예측 모델을 이용하여, 상기 타겟 시스템에 대한 장애를 예측하는 시스템 장애 예측 모듈을 포함하는 장애 예측 시스템이 제공된다.

발명의 효과

[0014] 본 발명의 일실시예에 따르면, 복수의 장애 예측 모델의 성능 평가 결과에 기반하여, 일부 장애 예측 모델을 선택적으로 재학습하여 장애를 예측함으로써, 장애 예측 모델의 학습 과정에 소요되는 비용과 시간이 줄어들 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일실시예에 따른 시스템의 장애를 예측하는 시스템을 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 시스템 장애 예측을 위한 학습 방법을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 시스템 장애 예측 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0017] 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.

[0019] 본 발명은, 현재 수집된 타겟 시스템에 대한 데이터를 이용하여, 미래에 타겟 시스템의 장애가 발생할지 여부를 예측하는 방법 및 시스템에 관한 발명이다.

[0020] 시스템 장애는 근본 원인(Fault)으로부터 시작된다. 근본 원인이 발생하면 시스템이 정상상태에서 벗어난 에러(Error)상태로 진입한다. 에러의 영향으로 시스템 상태의 이상 징후(Symptom)가 관측되며, 에러가 시스템의 비정상성을 유발하면, 장애(Failure)가 된다. 이와 같이, 시스템 장애의 발생 전에는 이상 징후가 관측되므로, 현재 수집된 데이터를 이용하여 미래에 발생할 장애가 예측될 수 있다.

[0021] 본 발명의 일실시예는, 기계 학습 기반의 장애 예측 모델을 이용하여, 시스템의 장애를 예측한다. 통신 시스템, 클라우드 시스템, 가상화 시스템 등 다양한 시스템이 존재하며, 동일한 시스템도 이용 환경에 따라서 장애 발생 여부가 달라지기 때문에, 본 발명의 일실시예는, 미리 학습된 복수의 장애 예측 모델을 이용하여 시스템의 장애를 예측한다. 복수의 장애 예측 모델 중에서, 타겟 시스템의 장애 예측 성능이 가장 우수하다고 평가되는 장애 예측 모델을 선택적으로 이용하여, 타겟 시스템의 장애를 예측한다.

[0022] 장애 예측 모델의 예측 성능을 높이기 위해서는 많은 훈련 데이터를 이용하여 학습 횟수를 높이는 것이 중요하며, 따라서 장애 예측 모델에 대한 재학습이 필요하다. 재학습에 소요되는 시간과 비용이 상당하기 때문에, 본 발명의 일실시예는, 재학습에 소요되는 시간과 비용을 줄이기 위해, 타겟 시스템의 장애 예측 성능이 우수하다고 평가되는 장애 예측 모델 후보군을 선정하여, 재학습을 수행한다.

[0023] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0025] 도 1은 본 발명의 일실시예에 따른 시스템의 장애를 예측하는 시스템을 설명하기 위한 도면이다.

- [0026] 도 1을 참조하면, 본 발명의 일실시예에 따른 장애 예측 시스템은, 데이터 수집 모듈(110), 데이터 관리 모듈(120), 모델 추천 모듈(130), 최적화 모듈(140) 및 시스템 장애 예측 모듈(150)을 포함한다. 모듈 각각은 용이한 배포를 위한, 컨테이너 기반의 모듈일 수 있다. 즉, 모듈 각각은 도커 컨테이너(docker container)와 같이, 컨테이너로 구현된 모듈일 수 있다.
- [0027] 데이터 수집 모듈(110)은 타겟 시스템에 대한 시스템 데이터를 수집한다. 일실시예로서, 데이터 수집 모듈(110)은 모나스카 에이전트(monasca agent)와 같은 에이전트를 통해 클라이언트의 시스템 데이터를 수집할 수 있으며, 사용자 설정 정보 등을 추가로 수집할 수 있다. 이 때, 수집되는 시스템 데이터는 사용자 설정 정보에 의해 결정될 수 있다. 이와 같이 수집된 시스템 데이터는, 훈련용 데이터로 이용될 뿐만 아니라, 장애 예측용 데이터로 이용될 수 있다.
- [0028] 시스템 데이터는 일실시예로서, 미리 설정된 수집 주기에 따라서 수집된 타겟 시스템의 CPU, 메모리, 디스크, 네트워크 등의 자원 정보일 수 있다. 예컨대, 특정 시간 T_n 에서 수집된 시스템 데이터 R_n 를, $T_n R_n$ 라고 표현하고, 특정 시간 T_n 에서의 시스템 장애 발생 여부를 Y_n 라고 표현할 경우, 데이터 수집 모듈(110)은 T_0 에서 T_{n-1} 까지
- $$\{T_0 R_0, T_0 R_1, T_0 R_2, \dots, T_0 R_{m-1}, Y_0\},$$
- $$\{T_1 R_0, T_1 R_1, T_1 R_2, \dots, T_1 R_{m-1}, Y_1\},$$
- $$\{T_{n-1} R_0, T_{n-1} R_1, T_{n-1} R_2, \dots, T_{n-1} R_{m-1}, Y_{n-1}\}$$
- 와 같이, 시계열 형태로 시스템 데이터를 수집할 수 있다.
- [0029] 데이터 관리 모듈(120)은 데이터 수집 모듈(110)에서 수집된 시스템 데이터를 전처리한다. 수집된 시스템 데이터에는, 활용하기 어려운 데이터들이 포함되어 있을 수 있으므로, 데이터 관리 모듈(120)은 수집된 시스템 데이터의 품질을 높이기 위해 전처리를 수행한다. 전처리 과정에서, 값이 불완전하거나 누락된 데이터, 노이즈 데이터 등이 제거된다.
- [0030] 그리고 데이터 관리 모듈(120)은 전처리된 데이터를 정규화하고 분산값을 조정하여 데이터 베이스에 저장할 수 있다. 전처리된 데이터는, 수집된 타겟 시스템 별로 분류되어 데이터 베이스에 저장될 수 있으며, 모나스카 플러그인(plugin) 별로 저장될 수 있다.
- [0031] 모델 추천 모듈(130)은 미리 학습된 복수의 장애 예측 모델에 대한 성능 평가값에 기반하여, 장애 예측 모델 중에서 재학습 장애 예측 모델 후보군을 결정한다. 즉, 모델 추천 모듈(130)은 복수의 장애 예측 모델 중에서 재학습을 위한 장애 예측 모델을 추천하며, 장애 예측 모델 후보군에 포함된 장애 예측 모델에 대해서만, 재학습이 이루어진다.
- [0032] 장애 예측 모델은 데이터 수집 모듈(110)에 의해 수집된 시스템 데이터를 통해 학습될 수 있으며, 성능 평가값으로 기계 학습 모델의 성능을 평가하는 다양한 평가값이 이용될 수 있다. 일예로서, 성능 평가값으로, F1-score, 중첩 교차 검증(nested cross validation)값 등이 이용될 수 있다.
- [0033] 모델 추천 모듈(130)은 실시예에 따라서, 장애 예측 모델에 대한 성능 평가값이 높은 순서로 장애 예측 모델에 대한 순위를 선정한 뒤, 미리 설정된 개수만큼의 상위 장애 예측 모델을 재학습 장애 예측 모델 후보군으로 선정할 수 있으며, 또는 성능 평가값이 임계값보다 큰 장애 예측 모델을 재학습 장애 예측 모델 후보군으로 선정할 수 있다.
- [0034] 최적화 모듈(140)은 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하고, 재학습된 장애 예측 모델에 대한 성능 평가값 중 최대값을 나타내는 베스트 장애 예측 모델을 결정한다. 또한 전술된 복수의 장애 예측 모델에 대한 학습을 수행할 수 있다.
- [0035] 최적화 모듈(140)은 학습 및 재학습 과정에서, 장애 예측 모델에 대해, 하이퍼 파라미터 튜닝이나, 시스템 데이터 중에서 학습을 위한 특징을 선택하는 등의 최적화를 수행한다. 그리고 재학습 및 최적화가 수행된 장애 예측 모델에 대한 성능을 평가하여, 베스트 장애 예측 모델을 데이터 베이스에 저장한다. 베스트 장애 예측 모델은, 성능 평가 결과가 가장 우수하여, 타겟 시스템에 대한 장애 예측이 가장 우수할 것으로 결정된 장애 예측 모델이다.
- [0036] 시스템 장애 예측 모듈(150)은 시스템 데이터 및 베스트 장애 예측 모델을 이용하여, 타겟 시스템에 대한 장애

를 예측한다.

- [0038] 도 2는 본 발명의 일실시예에 따른 시스템 장애 예측을 위한 학습 방법을 설명하기 위한 도면이다.
- [0039] 도 2를 참조하면, 본 발명의 일실시예에 따른 장애 예측 시스템은 타겟 시스템에 대한 시스템 데이터를 포함하는 제1훈련 데이터를 이용하여 복수의 장애 예측 모델을 학습(S210)한다. 그리고 장애 예측 모델에 대한 성능 평가값에 기반하여, 재학습 장애 예측 모델 후보군을 결정(S220)한다.
- [0040] 이 때, 단계 S220에서 장애 예측 시스템은, 재학습에 이용되는 제2훈련 데이터와 단계 S210에서 이용된 제1훈련 데이터의 패턴 유사도에 따라서, 복수의 장애 예측 모델 중에서 재학습 장애 예측 모델 후보군을 결정한다. 제2훈련 데이터는 제1훈련 데이터와는 다른 시점에 수집된 훈련 데이터일 수 있다.
- [0041] 장애 예측 시스템은 제1 및 제2훈련 데이터의 패턴 유사도를 판단하고, 제1 및 제2훈련 데이터의 패턴 유사도가 임계값 이상인 경우, 복수의 장애 예측 모델에서 재학습 장애 예측 모델 후보군을 결정할 수 있다. 예컨대, 제1 및 제2훈련 데이터에 포함된 자원 정보의 종류가 유사하고, 이러한 자원 정보가 시간 흐름에 따라 유사한 패턴을 나타낸다면, 장애 예측 시스템은, 복수의 장애 예측 모델에서 재학습 장애 예측 모델 후보군을 결정할 수 있다. 장애 예측 시스템은 데이터 패턴의 유사도를 판단하는 다양한 알고리즘을 이용할 수 있다.
- [0042] 본 발명의 일실시예에 따른 장애 예측 시스템은, 제2훈련 데이터를 이용하여, 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습(S230)한다. 제1 및 제2훈련 데이터의 패턴이 유사하지 않을 경우, 장애 예측 시스템은, 제1훈련 데이터를 통해 학습된 장애 예측 모델 전체를 제2훈련 데이터를 이용하여 재학습할 수 있다.
- [0043] 단계 S230에서 장애 예측 시스템은, 전처리된 제2훈련 데이터를 이용하여 재학습 장애 예측 모델 후보군에 포함된 장애 예측 모델을 재학습하며, 재학습된 장애 예측 모델 중 성능 평가값이 최대값을 나타내는 장애 예측 모델을, 베스트 장애 예측 모델을 결정하여 저장한다. 복수의 장애 예측 모델 중에서 베스트 장애 예측 모델이, 시스템의 장애 예측에 이용된다.
- [0044] 본 발명의 일실시예에 따르면, 복수의 장애 예측 모델 전체에 대해 반복적으로 재학습하지 않고, 타겟 시스템에 대한 장애 예측 성능이 우수할 것으로 평가되는 장애 예측 모델을 선택적으로 재학습함으로써 재학습에 소요되는 비용과 시간이 줄어들 수 있다.
- [0046] 도 3은 본 발명의 일실시예에 따른 시스템 장애 예측 방법을 설명하기 위한 도면이다.
- [0047] 본 발명의 일실시예에 따른 장애 예측 시스템은, 도 2에서 결정된 베스트 장애 예측 모델을 이용하여 타겟 시스템의 장애 발생을 예측한다.
- [0048] 장애 예측 시스템은, 타겟 시스템에 대한 타겟 시스템 데이터를 수집(S310)하고, 타겟 시스템 데이터 및 베스트 장애 예측 모델을 이용하여, 타겟 시스템의 장애 발생 여부를 예측(S320)한다. 타겟 시스템 데이터가 베스트 장애 예측 모델로 입력되면, 베스트 장애 예측 모델은 타겟 시스템에 대한 장애가 발생할 것인지, 아니면 발생하지 않을 것인지 여부를 출력한다. 예컨대, 베스트 장애 예측 모델은 장애가 발생할 것으로 예측한 경우 1, 장애가 발생하지 않을 것으로 예측한 경우 0을 출력할 수 있다.
- [0049] 타겟 시스템에 대한 장애가 발생할 것으로 예측된 경우, 장애 예측 시스템은 베스트 장애 예측 모델의 성능 평가값을 이용하여, 베스트 장애 예측 모델의 장애 발생 예측 정확도를 제공(S330)한다. 장애 발생 예측 정확도는, 베스트 장애 예측 모델의 장애 예측 결과가 얼마나 정확한지를 나타내는 값으로서, 베스트 장애 예측 모델의 성능 평가값에 따라 결정된다.
- [0050] 베스트 장애 예측 모델의 성능 평가값이 높으면 장애 발생 예측 정확도가 높은 것으로 해석될 수 있으며, 베스트 장애 예측 모델의 성능 평가값이 낮으면 장애 발생 예측 정확도 역시 낮은 것으로 해석될 수 있다. 베스트 장애 예측 모델의 성능 평가값이 높을수록 베스트 장애 예측 모델의 예측 결과에 대한 신뢰도가 높아진다는 점에서, 장애 발생 예측 정확도는 장애 발생 예측 신뢰도로 표현될 수도 있다.
- [0051] 본 발명의 일실시예에 따르면, 베스트 장애 예측 모델의 장애 예측 결과와, 장애 발생 예측 정확도가 함께 제공됨으로써, 관리자의 시스템 장애 대응을 효율적으로 지원할 수 있다.
- [0052] 한편, 실시예에 따라서, 장애 발생 예측 정확도는 베스트 장애 예측 모델의 최적화 횟수 및 장애 예측 이력 정보 중 적어도 하나에 따라서 조절될 수 있다.
- [0053] 장애 예측 모델에 대한 성능 평가값은 장애 예측 모델에 대한 최적화 이후 도출될 수 있으며, 복수의 최적화가

이루어질 경우 각각의 최적화 이후 성능 평가값이 도출될 수 있는데, 장애 예측 시스템은 베스트 장애 예측 모델의 성능 평가값의 합을 최적화 횟수로 나눈 평균값을 장애 발생 예측 정확도로 제공할 수 있다.

[0054] 또는 장애 예측 시스템은, 장애 예측 이력 정보를 이용하여, 장애 발생 예측 정확도를 계산하여 제공할 수 있다. 장애 예측 이력 정보는 베스트 장애 예측 모델의 장애 발생 예측 내용과, 타겟 시스템의 장애 발생 사이의 일치 비율을 포함한다. 즉, 베스트 장애 예측 모델이 과거에 타겟 시스템의 장애 발생이 발생할 것으로 예측한 결과 중에서, 실제 장애 발생으로 이어진 사례의 비율 정보를 포함한다.

[0055] 또한 장애 예측 이력 정보는 베스트 장애 예측 모델의 장애 발생 예측 시점과 타겟 시스템의 장애 발생 시점 사이의 시간 차이값을 포함할 수 있다. 즉, 베스트 장애 예측 모델이 장애가 발생할 것으로 예측한 결과가 제공된 시점으로부터 실제 장애가 발생한 시점 사이의 시간 차이값을 포함할 수 있다.

[0056] 장애 발생 예측 정확도는 일치 비율 및 시간 차이값에 비례할 수 있다. 시스템의 장애 발생 시점에 임박할수록 시스템의 장애 징후는 증가하며, 따라서, 장애가 발생할 것으로 예측한 결과가 제공된 시점으로부터 실제 장애가 발생한 시점 사이의 시간 차이값이 크다는 것은, 적은 장애 징후에도 정확히 장애를 예측한 것으로 볼 수 있으므로, 장애 예측 시스템은 시간 차이값에 비례하도록 장애 발생 예측 정확도를 계산하여 제공할 수 있다.

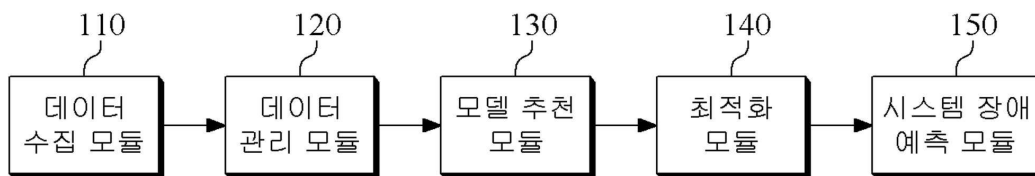
[0057] 장애 예측 시스템은, 베스트 장애 예측 모델의 성능 평가값에 대한 평균값을 장애 예측 이력 정보에 따라 조절하여, 장애 발생 예측 정확도로 제공할 수 있다.

[0059] 앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

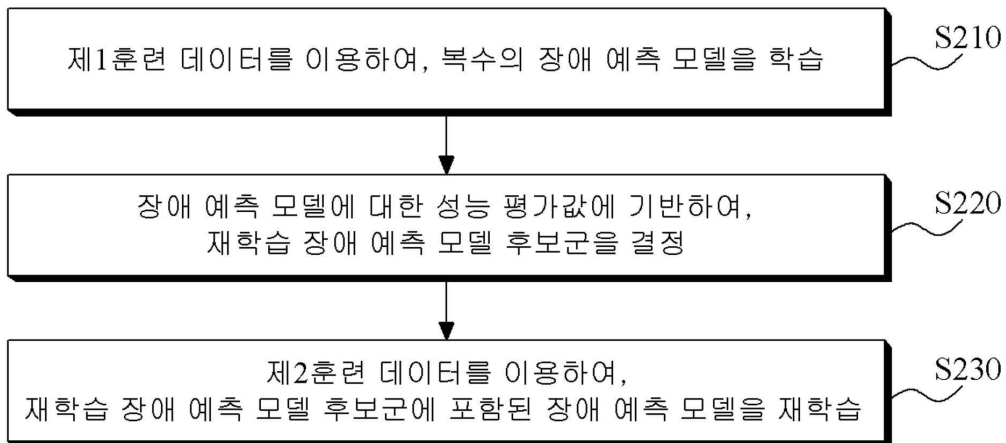
[0061] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

도면1



도면2



도면3

