

논문 2026-63-2-5

다변량 시계열 이상 탐지를 통한 블록체인 검증자 신뢰성 확보

(Ensuring Blockchain Validator Reliability through Multivariate Time-series Anomaly Detection)

이 승 재*, 김 명 선**

(Seungjae Lee and Myungsun Kim[©])

요 약

블록체인 검증자는 블록체인 네트워크의 가용성과 무결성을 보장하는 데 매우 중요하다. 하지만 검증자의 신뢰성을 유지하기 위한 연구는 아직 활발하지는 않다. 본 논문은 블록체인 검증자를 위해 특별히 설계된 딥러닝 기반 이상 탐지 프레임워크를 제안하며, 이는 검증자의 고차원 원격 측정 데이터로 인해 발생하는 문제를 해결한다. '차원의 저주'를 극복하기 위해 이상 탐지 전처리 단계로 차원 축소 기법을 적용한다. 우리는 여러 차원 축소 기법(PCA, UMAP, LLE, ISOMAP)과 이상 탐지 모델(LSTM-VAE, VTTSAT, VTTPAT)의 조합을 체계적으로 평가한다. 실험 결과, PCA와 VTTSAT의 조합이 0.7922의 PA%K AUC를 기록하며 최고의 성능을 보였다. 더 나아가, 축소된 차원 수를 변경하며 진행한 실험에서 50차원이 검증자 이상 탐지에 가장 효과적인 표현을 제공한다는 것을 발견했다. 우리의 연구 결과는 딥러닝 기반 이상 탐지를 검증자 환경에 적용하는 것의 실현 가능성을 강조하며, 이를 통해 블록체인 운영의 신뢰성을 향상시킬 수 있음을 보여준다.

Abstract

Blockchain validators are crucial for ensuring the availability and integrity of blockchain networks. However, research on maintaining validator reliability has not yet been active. This paper proposes a deep learning-based anomaly detection framework specifically designed for blockchain validators, addressing challenges arising from their high-dimensional telemetry data. To overcome the 'curse of dimensionality,' dimensionality reduction techniques are applied as a preprocessing step for anomaly detection. We systematically evaluate combinations of various dimension reduction techniques (PCA, UMAP, LLE, ISOMAP) and anomaly detection models (LSTM-VAE, VTTSAT, VTTPAT). Experimental results show that the combination of PCA and VTTSAT achieved the best performance with a PA%K AUC of 0.7922. Furthermore, experiments varying the reduced dimension count revealed that 50 dimensions provided the most effective representation for validator anomaly detection. Our findings highlight the feasibility of applying deep learning-based anomaly detection in validator environments, demonstrating its potential to enhance the reliability of blockchain operations.

Keywords : Anomaly detection, Dimension reduction, Blockchain validator

I. 서 론

블록체인은 탈중앙화된 분산 원장 관리 기술 중 하나로, 중앙화된 정보 관리와 달리 데이터를 분산시켜 저장하는 기술이다^[1]. 사용자가 기록한 데이터들이 모여

하나의 블록 형태로 블록체인에 저장된다. 사용자는 수행하는 거래, 정보 데이터를 블록체인에 기록하고, 블록체인에 저장한 데이터를 조회할 수 있다. 이전 블록이 다음 블록에 포함되는 형태이기에 생성된 블록은 조작, 변경되는 것이 불가능하다.

이러한 특성으로 인해 블록체인은 여러 분야에서 광범위하게 사용되는 기술이 되었다. 최근 정부는 블록체인에 기반한 탈중앙화된 신원 증명 플랫폼을 도입했다^[2]. 또한 물류 업계에서도 물류 이력 관리를 위해 블록체인을 도입하고 있다^[3].

*학생회원, **정회원, 한성대학교 AI융용학과
(Department of Applied AI, Hansung University)

© Corresponding Author(E-mail : kmsjames@hansung.ac.kr)

※ 본 연구는 한성대학교 교내학술연구비 지원과제임.

Received ; September 16, 2025 Revised ; September 22, 2025

Accepted ; September 23, 2025

블록체인을 운영하기 위해서는 검증자라는 노드가 필수적이다. 검증자는 블록체인의 무결성, 가용성을 보장하는 역할을 한다. 따라서 블록체인 검증자는 최대한 활성 상태를 유지해야 하며, 검증자가 비활성 상태, 이상 상태라면 최대한 빨리 이를 감지하고 조치하는 것이 필요하다.

검증자의 원격 측정 데이터는 본질적으로 고차원적이며, 시스템 수준 변수(CPU 사용량, 메모리, 네트워크 통계)와 블록체인 특정 기능(피어 수, 블록 전파 지연 시간, 검증자 참여, 트랜잭션 수)을 포함하는 다변량 시계열 데이터이다. 이러한 데이터 환경에서의 이상 탐지는 다변량 시계열 데이터에서 이상 시점을 찾아내는 것을 뜻한다. 다변량 시계열 데이터는 일정한 시간 간격이 있는 여러 개의 시점으로 구성되는 데이터이다. 즉, 다변량 시계열 데이터에서 각 시점이 이상인지, 정상인지 분류하는 작업이 다변량 시계열 이상 탐지이다^[4].

늦지 않게 이상을 탐지하는 것은 가용성이 중요한 산업에서 필수적이다. 제조업, 클라우드 등 가용성이 중요한 산업들은 딥러닝 기반 이상 탐지 모델을 만들고, 이를 적극적으로 사용해 생산성을 높이려 한다^[5]. 다만 딥러닝 기반 이상 탐지는 어려운 작업이다. 실제 산업 현장에서 이상 데이터를 확보하기 어렵고, 발생할 수 있는 이상 패턴이 다양하기 때문이다.

이러한 이유로 이상 탐지 모델의 학습 데이터는 정상 데이터 비율이 매우 높거나, 정상 데이터로만 이루어진다. 이러한 클래스 불균형 때문에, 대부분의 이상 탐지 모델은 비지도 학습을 사용한다. 이상 탐지는 이상 데이터가 들어왔을 때 모델이 학습했던 정상 데이터와 다름을 감지하는 방식으로 탐지한다^[6].

본 논문에서는 블록체인 운영에 필수적인 검증자의 활성 상태 유지를 위해 검증자 전용 딥러닝 기반 이상 탐지 방법론을 제안한다. 검증자 환경을 구축하고 데이터를 수집한다. 검증자의 복잡한 상태를 나타내는 고차원 데이터를 모델 학습에 활용하기 위해 차원 축소 기법으로 전처리한다. 변환된 저차원 데이터를 이상 탐지 모델의 학습, 테스트 데이터로써 활용한다.

II. 연구배경 및 문제정의

1. 블록체인 검증자

하나의 블록체인을 운영하기 위해서는 검증자 노드가 필요하다. 검증자는 다음과 같은 기능을 수행한다. 1) 검증자는 블록체인에 제출된 데이터를 취합한다. 2) 검

증자는 1)의 데이터를 받아 블록을 생성한다. 3) 검증자는 다른 검증자가 만든 블록이 올바른지 검증한다^[7]. 즉, 검증자는 전체 블록체인의 무결성을 보장하는 노드로써, 블록체인 운영에서 핵심적인 역할을 한다. 블록체인의 원활한 운영은 검증자의 정상적인 기능 수행이 보장되어야만 한다^[7, 8].

활성 상태를 유지하는 검증자는 블록체인 운영에 필수적임에도 불구하고, 검증자 운영, 활동 유지, 모니터링 연구는 많이 이루어지지 않고 있다. 본 논문에서는 블록체인의 원활한 운영을 위한 검증자 이상 탐지 방법론을 제안하고 실험한다.

2. 다변량 시계열 이상 탐지

다변량 시계열 데이터 이상 탐지 모델은 크게 2가지 분류가 있다. 예측 기반 이상 탐지 모델은 이전 시점의 데이터를 받아 현재 시점의 데이터를 예측한다. 만일 모델의 예측한 시점의 데이터와 실제 시점의 데이터가 크게 다르다면, 그 시점을 이상이라고 분류한다. 이러한 예측 기반 이상 탐지 모델은 LSTM-AD^[9], DeepAnT^[10], FuseAD^[11] 등이 있다.

재구성 기반 이상 탐지 모델은 입력 데이터를 잠재 공간으로 인코딩하고 다시 입력 데이터로 재구성한다. 입력 데이터와 재구성된 데이터의 차이를 시점별로 구하고, 만약 차이가 크다면 해당 시점을 이상으로 분류한다. 이러한 재구성 기반 이상 탐지 모델은 LSTM-VAE^[12], VTTSAT, VTTPAT^[13] 등이 있다.

본 논문에서는 딥러닝 기반 이상 탐지 모델을 사용해 검증자의 이상을 탐지한다.

3. 문제정의

딥러닝 기반 이상 탐지를 검증자에 적용하기엔 어려움이 있다. 첫 번째로, 검증자 이상 탐지 모델을 학습하고 테스트하기 위한 검증자 데이터가 없다. 두 번째로, 검증자의 복잡한 상태를 나타내기 위해 수집해야 할 변수가 너무 많다. 즉, 검증자에서 수집되는 데이터가 고차원 데이터이다. 학습에 고차원 데이터가 사용되면, 모델이 과악해야 할 변수가 너무 많아 모델의 학습이 어렵다^[14]. 따라서 검증자 이상 탐지를 수행하기 위해서는 검증자 데이터 수집하는 과정, 고차원 데이터를 이상 탐지 모델에 사용하기 위해서 차원을 줄이거나 통합하는 과정이 필요하다.

첫 번째 문제를 해결하기 위해, 실제 이더리움 환경에서 동작하는 검증자를 구축한다. 모델 학습을 위한

정상 데이터와 모델 테스트를 위한 이상 데이터를 수집한다. 이후 두 번째 문제를 해결하기 위해 수집한 고차원 데이터의 차원을 줄이거나 통합한다. 고차원 데이터에 대한 처리는 차원 축소로 수행한다. 차원 축소는 고차원 데이터를 저차원 데이터로 변환하는 것이다^[15]. 대표적인 차원 축소 기법에는 PCA^[16], UMAP^[17], LLE^[18], ISOMAP^[19] 등이 있다.

III. 차원 축소 및 딥러닝을 활용한 블록체인 검증자 이상 탐지 기법

본 논문의 목적은 블록체인에서 핵심적인 기능을 수행하는 검증자의 가용성 유지를 위해, 블록체인 검증자의 이상 탐지에 활용할 수 있는 딥러닝 기반 이상 탐지 모델을 학습하고, 실제 환경에서 활용할 수 있는지 확인하는 것이다.

1. 검증자 데이터 수집

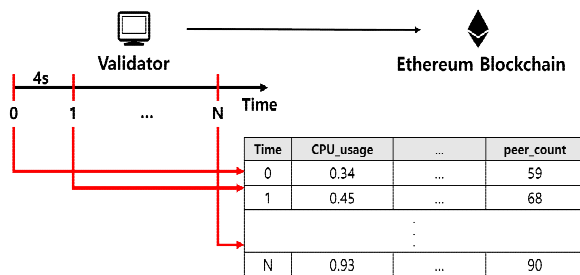


그림 1. 검증자 데이터 수집 과정 모식도
Fig. 1. Schematic illustration of the validator data collection process.

그림 1은 검증자 데이터 수집 과정의 모식도이다. 이더리움 블록체인상에서 동작하는 검증자에서 데이터를 수집했다. 다변량 시계열 데이터의 수집은 여러 변수를 일정한 시간 간격이 있는 시점을 기록하는 방식으로 이뤄진다.

시점별로 1행의 데이터를 수집했다. 시점 간의 간격은 4초이다. 시점 간의 간격을 4초로 설정한 이유는, 이더리움 블록체인에서 검증자가 작동하는 시간 단위인 slot이 12초이기 때문이다^[20]. 1개 slot당 3개의 행을 수집하기 위해, 4초 간격을 두어 데이터를 수집했다.

총 746개의 변수가 수집되었으며, 변수 항목은 CPU 사용량, 네트워크 패킷, 메모리 사용량 같은 시스템 변수와 블록체인 상태를 나타내는 P2P 피어 수, 최신 블

록 수신 지연시간, 검증 참여 횟수, 블록체인 동기화 여부, 블록에 포함되는 트랜잭션 수와 같은 블록체인 관련 변수로 분류된다.

각 시점마다 수집된 데이터는 각 시점의 검증자 시스템의 상태를 표현한다.

표 1. 검증자 명세

Table 1. Validator System specifications.

System	Classification	Description
HW	CPU	13th Gen Intel(R) Core(TM) i7-13700
	Memory	2 x Samsung M378A4G43AB2-CWE 32GB
	Storage	Samsung MZVL22T0HBLB-00B00 SSD 2TB
SW	Execution Layer	Besu - 25.3.0
	Beacon Node	Lighthouse - v7.0.0-beta.4-0486802
	Validator Client	Lighthouse - v7.0.0-beta.4-0486802

검증자 프로그램이 실행된 환경은 표 1과 같다. 위 환경에서 1개의 검증자를 실행했고, 데이터 수집은 31 시간 동안 진행되었다.

표 2. 검증자 데이터 명세

Table 2. Validator data specifications.

Details	Validator Dataset
Variables	746
Number of Entity	1
Train	10000
Test(labeled)	17500
Anomaly(%)	36.16%

표 2는 수집한 검증자 데이터에 대한 대략적인 명세이다. 746개의 변수가 4초 간격으로 27,500 시점(행) 수집되었다. 학습 데이터는 10,000행으로 구성되어 있다. 일반적인 이상 탐지 방법론을 적용하기 위해, 학습 데이터는 정상 동작하는 검증자에게서만 수집되었다.

테스트 데이터는 17,500행으로 구성되어 있다. 검증자를 실제 운영할 때 발생하는 문제를 고려해, 총 7가지의 이상 상황을 가정했다. ①, ②, ③ 검증자 역할을 수행하기 위한 3개의 프로그램, Execution Layer, Beacon Node, Validator Client를 각각 종료된 상황,

④ 검증자 시스템의 네트워크가 차단된 상황, ⑤, ⑥, ⑦ Execution Layer, Beacon Node, Validator Client의 데이터베이스가 삭제된 상황을 가정했다. 모든 이상 상황은 테스트 데이터를 수집하는 동안 주어졌고, 각 이상 상황은 1시간 동안 지속되었다. 이상 상황이 끝나고, 정상인 상태가 1시간 30분 동안 지속되었다. 이는 새로운 이상 상황이 발생하기 전, 이전에 발생했던 이상 상황의 영향을 배제하기 위함이다.

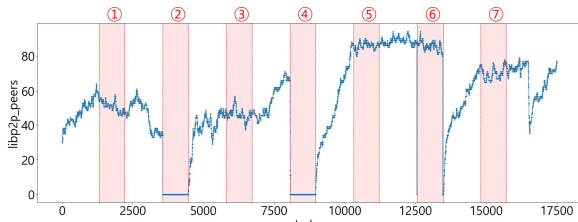


그림 2. 테스트 데이터 중 피어 수 항목 예시
Fig. 2. Example of number of peers in the test data.

그림 2는 테스트 데이터의 블록체인 관련 변수 중 하나인 p2p 피어 수 항목을 시각화한 것이다. 검증자가 원활히 동작하기 위해서는, 충분히 많은 수의 피어와 연결되어야 한다. x축은 행 번호, y축은 p2p 피어 수를 나타낸다. 또한 그래프 위 ①~⑦은 각 이상 상황을 의미한다. ② Beacon Node가 종료되는 이상 상황, ④ 네트워크가 종료되는 이상 상황에서 피어 수가 급격히 감소하는 이상 징후가 나타나는 것을 알 수 있다.

2. 이상 탐지

수집한 검증자 데이터는 746개의 변수를 가진 고차원 데이터이다. 고차원 데이터를 학습에 사용하면, 모델이 파악해야 하는 변수 간의 관계, 변수의 패턴이 너무 많아 학습이 진행되지 않는다. 따라서 여러 차원을 통합하거나, 차원을 없애는 과정이 필수적이다.

그림 3은 검증자 이상 탐지 방법론 흐름도이다. 차원 축소를 통해 N 차원을 N' 차원으로 줄인다. 이후 축소된 저차원 데이터는 이상 탐지 모델의 입력으로 사용된다. VTTSAT 모델 예시에선, 내부적으로 시점 간 관계, 변수 간 관계를 파악한다. 이후 출력으로 나온 재구성 값을 입력 값과 비교해 이상을 탐지한다.

차원 축소를 이상 탐지 모델의 학습, 테스트 전에 수행한다. 본 논문에서는 차원 축소 기법을 이용해 746개의 변수를 가진 고차원 데이터를 저차원 데이터로 변환한다. 저차원 데이터를 이상 탐지 모델의 학습, 테스트

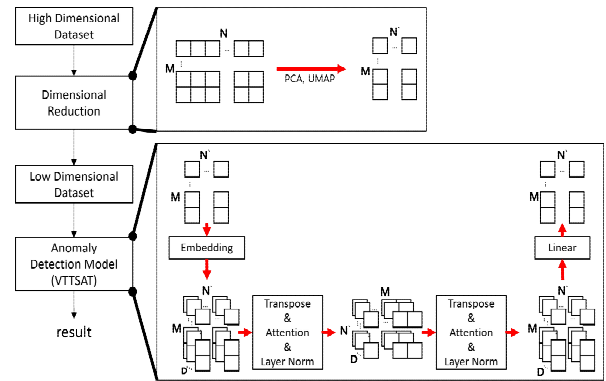


그림 3. 검증자 이상 탐지의 동작 흐름도
Fig. 3. Workflow of anomaly detection in a validator.

데이터로 활용한다.

이후 다음 장에서 고차원 데이터, 저차원 데이터를 학습에 사용한 결과, 4개의 차원 축소 기법을 전처리에 달리 적용했을 때의 결과, 전처리의 결과인 저차원 데이터의 차원을 달리했을 때의 결과를 각각 비교, 분석한다.

IV. 실험

본 장에서는 LSTM-VAE를 이용해, 검증자에서 수집한 고차원 데이터를 어떠한 전처리 없이 이상 탐지에 그대로 사용한 결과와 차원 축소 전처리를 적용한 저차원 데이터를 이상 탐지에 사용한 결과를 비교하는 실험을 진행한다. 이는 고차원 데이터를 실제로 이상 탐지 모델에 사용할 수 없음을 보이기 위함이다.

이상 탐지에 있어 차원 축소 기법의 성능을 비교하기 위해, 검증자 데이터의 전처리 과정에서 사용한 차원 축소 기법을 달리해가며 결과를 비교하는 실험을 진행한다. 이때 이상 탐지 모델의 영향을 줄이기 위해 VTTSAT, VTTPAT 2개 모델 결과를 종합해서 분석한다. 또한 차원 축소 기법의 결과인 저차원 데이터의 차원 수를 달리해서, 어떤 저차원 데이터의 차원 수가 검증자 이상 탐지에 제일 적합한지 탐색하는 실험을 진행한다.

1. 성능 지표

다변량 시계열 이상 탐지 모델에서 성능을 확인하는 지표는 2가지가 있다. $F1_{PA}$ score는 이상 시점이 연속해서 나타나는 실제 환경의 특징에 맞춰 만들어진 지표이다. 이상 시점이 연속해서 나타나는 구간을 이상 구간이라고 한다. 이상 구간 내의 하나의 시점만을 이상

으로 탐지한 상황에서, $F1_{PA}$ score를 계산할 때, 이상 구간 내의 모든 시점을 옳게 탐지한 것으로 간주한다. 반면 동일한 상황에서 $F1$ score는 이상 구간 내에서 탐지한 하나의 시점만을 옳게 분류한 것으로 처리해 산출한다.

$F1_{PA}$ score는 이상 탐지 모델의 성능을 과대평가하고, $F1$ score는 실제 환경의 특징을 전혀 반영하지 못한다는 문제가 있다. 이를 해결하기 위해 $F1_{PA\%K}$ score라는 지표가 사용된다^[21]. $F1_{PA\%K}$ score는 K에 따라 지표의 산출이 달라진다. K는 일종의 임계값이며, 이상 구간 내에서 K%의 이상 시점을 탐지했을 때, 해당 이상 구간의 모든 시점을 탐지한 것으로 간주한다. 즉, $K = 0$ 일 때 $F1_{PA}$ score이고, $K = 100$ 일 때, $F1$ score이다.

따라서 K가 변함에 따라, 여러 $F1_{PA\%K}$ score가 나오게 된다. K가 x축이고, $F1_{PA\%K}$ score이 y축인 도표에 각 점을 나타내고, 근접한 K의 두 점끼리 이으면 1개의 곡선이 나타나게 된다.

이 곡선의 이름을 $F1_{PA\%K}$ curve라 하고, 이 곡선의 면적을 PA%K AUC라 한다^[21]. PA%K AUC는 모든 K에 대해 각 $F1_{PA\%K}$ score를 종합한 지표이다. 이후 실험은 $F1_{PA\%K}$ score와 PA%K AUC를 지표로 한다.

2. 실험 환경

표 3. 실험 환경 명세

Table 3. Validator system specification.

System	Classification	Description
HW	CPU	AMD Ryzen Threadripper PRO 3955WX 16-Cores
	GPU	NVIDIA RTX A6000 D6 48GB
	Memory	4 x Samsung M393A8G40AB2-CWE 64GB
	Storage	Samsung - MZ7LH1T9HMLT 1.92TB
SW	CUDA Ver.	CUDA v12.2
	Pytorch Ver.	2.4.0+ cu121
	Kernel Ver.	Linux 5.15.0

표 3은 차원 축소 전처리와 이상 탐지 모델 학습 및 테스트를 진행한 환경의 사양이다. 실험에 사용되는 모든 모델은 100 epoch 동안 학습했고, 배치 사이즈는 32로 수행했다. 실험 지표는 K를 0, 1, 2, 3, 4, 5, 10,

20, 30, 40, 50, 60, 70, 80, 90, 100으로 설정해 각 $F1_{PA\%K}$ score를 측정했고, 이를 기반으로 PA%K AUC를 산출했다.

3. 고차원 데이터와 저차원 데이터 사용 시 비교

이 실험에서는 다차원 데이터에 전처리를 가하지 않고 이상 탐지를 한 경우와 다차원 데이터에 차원 축소 전처리를 수행해 변환된 저차원 데이터로 이상 탐지를 한 경우를 비교한다.

실험에 사용한 모델은 LSTM-VAE이고, VAE의 잠재 공간은 3으로 설정했다^[12]. 저차원 데이터의 차원 수는 50으로 설정했고, 차원 축소 기법은 PCA를 사용했다.

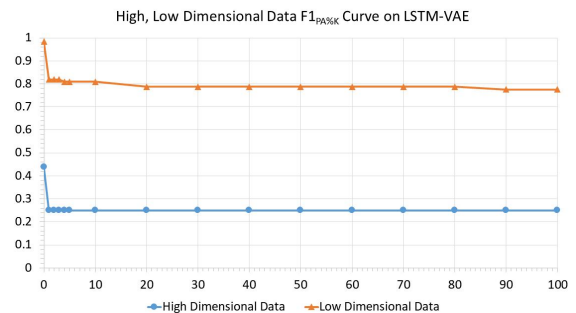


그림 4. 고차원 데이터, 저차원 데이터 결과 비교

Fig. 4. Comparison of high dimensional data with low dimensional data results.

표 4. 고차원 및 저차원 데이터 PA%K AUC 비교

Table 4. Comparison of PA%K AUC for high dimensional and low dimensional data.

Data	PA%K AUC
High dimensional data	0.2508
Low dimensional data	0.7902

그림 4는 LSTM-VAE 모델로, 저차원 데이터를 통한 이상 탐지, 고차원 데이터를 통한 이상 탐지를 수행했을 때의 $F1_{PA\%K}$ score curve이다. 표 4는 그림 4의 $F1_{PA\%K}$ score curve로부터 PA%K AUC를 산출한 결과이다. 고차원 데이터의 PA%K AUC는 0.2508로 상당히 낮은 것을 알 수 있다. 그 반면, 저차원 데이터의 PA%K AUC는 0.7902로, 고차원 데이터의 결과와 0.5394의 차이가 난다. 이러한 차이는 모델의 학습이 고차원 데이터로 진행될 경우, 변수가 너무 많아 유의미한 변수 간의 관계, 패턴의 학습이 제대로 되지 않았음을 보여준다.

위 실험은 고차원 데이터를 검증자 이상 탐지에 바로 활용할 수 없다는 것과 고차원 데이터를 저차원으로 변환하는 차원 축소 전처리가 필요함을 보여준다.

4. 차원 축소 기법 간 비교

이 실험에서는 검증자 이상 탐지에서 어떤 차원 축소 기법 전처리가 적합한지 판단한다. 전처리 기법을 달리 해 이상 탐지를 진행하고, 결과를 비교한다. 차원 축소 기법으로는 ISOMAP, LLE, PCA, UMAP을 사용한다.

추가로 차원 축소 기법을 이용한 전처리가 사용자가 직접 수행하는 차원 축소에 비해 고려할 만한 선택지임을 보이기 위해, 사용자가 차원을 축소한 데이터로 모델 학습, 테스트를 진행하고, 이 테스트 결과와 함께 비교한다. 사용자가 수행하는 차원 축소의 기준은 검증자의 상태를 잘 나타낼 수 있도록, 시스템 변수와 블록체인 변수의 비율을 절반으로 조절해 선택했다.

사용자가 수행한 차원 축소를 포함해, 모든 차원 축소는 746차원의 데이터를 50차원으로 변환했다. 실험에 사용된 모델은 VTTSAT와 VTTPAT이다.

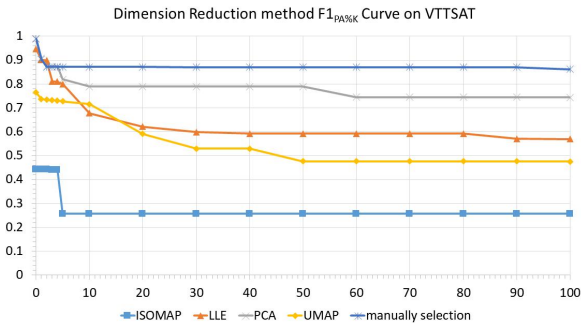


그림 5. VTTSAT에서의 차원 축소 기법 비교
Fig. 5. Dimension reduction method compare on VTTSAT.

표 5. VTTSAT에서의 차원 축소 기법 비교
Table 5. Dimension Reduction method compare on VTTSAT.

Method	PA%K AUC
ISOMAP	0.2660
LLE	0.6155
PCA	0.7922
UMAP	0.5724
manually selection	0.8707

그림 5는 VTTSAT에서 각 차원 축소 기법 별 $F1_{PA\%K}$ score curve를 나타낸 결과이다. 표 5는 각 차원 축소 기법의 PA%K AUC를 나타낸다. 실험 결과,

차원 축소 기법에서는 PCA가 0.7922로, 제일 높은 PA%K AUC를 보여준다. ISOMAP을 사용한 테스트 결과는 0.2660으로, 차원 축소 기법 중 제일 낮은 PA%K AUC를 보여준다.

사용자가 직접 차원 축소를 수행했을 때의 PA%K AUC는 0.8707로 PCA의 AUC와 비교했을 때, 차이는 0.0785로 작은 차이를 보인다.

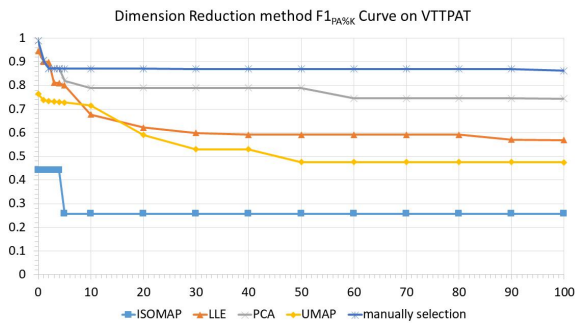


그림 6. VTTPAT에서의 차원 축소 기법 비교
Fig. 6. Dimension reduction method compare on VTTPAT.

표 6. VTTPAT에서 차원 축소 기법 비교
Table 6. Dimension reduction method compare on VTTPAT.

Method	PA%K AUC
ISOMAP	0.2657
LLE	0.6174
PCA	0.7750
UMAP	0.5356
manually selection	0.8710

그림 6은 VTTPAT에서 각 차원 축소 기법별로 $F1_{PA\%K}$ score curve를 나타낸 결과이다. 그림 5와 추세가 일치함을 알 수 있다. 표 6은 그림 6을 기반으로 계산된 PA%K AUC를 나타낸다. VTTPAT에서도 PCA가 제일 높은 PA%K AUC인 0.7750을 보여준다. 제일 낮은 PA%K AUC는 ISOMAP으로 0.2657이다.

VTTPAT에서도, 사용자가 차원 축소를 수행했을 때의 AUC가 제일 높음을 알 수 있다. 이때의 차이는 0.096이다.

두 실험을 종합하면, 두 모델에서 제일 좋은 성능을 보이는 차원 축소 기법은 PCA이며, 사용자가 직접 차원 축소를 수행하는 것이 제일 좋으나, 그럴 수 없는 경우 PCA를 차원 축소 전처리로 사용하는 것이 유효한 선택지임을 보여준다. 또한, 차원 축소 전처리 기법과 딥러닝 기반 이상 탐지를 블록체인 검증자 환경에서 활

용할 수 있음을 확인하였다.

5. 저차원 데이터 차원 수 간 비교

본 실험에서는 검증자 이상 탐지에서 저차원 데이터의 차원 수가 이상 탐지에 미치는 영향을 알아보고, 이상 탐지에서 제일 적절한 차원 수를 알아본다.

모든 저차원 데이터에서 PCA가 차원 축소 전처리로 사용되었다. 각 저차원 데이터의 차원 수는 50, 25, 10 이고, 각 저차원 데이터로 이상 탐지 모델을 학습하고 테스트했을 때의 결과를 비교한다.

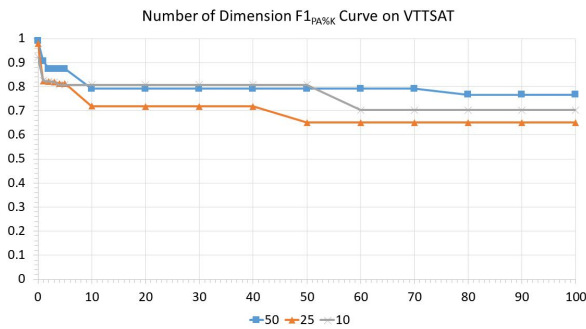


그림 7. 저차원 데이터의 차원 수 별 비교(VTTSAT)

Fig. 7. Comparison of low dimensional data by number of dimensions (VTTSAT).

표 7. 저차원 데이터의 차원 수 별 PA%K AUC (VTTSAT)

Table 7. PA%K AUC of low dimensional data by number of dimensions (VTTSAT).

Method	PA%K AUC
50	0.7922
25	0.6898
10	0.7617

그림 7은 차원 수가 50, 25, 10인 저차원 데이터에서 VTTSAT를 통한 이상 탐지를 수행했을 때의 F1PA%K score curve이다. 표 7은, 그림 7을 기반으로 계산된 PA%K AUC이다. 실험 결과, 저차원 데이터의 차원 수가 50일 때, PA%K AUC는 0.7922로 제일 높았고, 25일 때, 0.6898로 제일 낮았다.

V. 결 론

본 논문에서는 블록체인 운영에 필수적인 검증자에서의 이상 탐지를 수행한다. 블록체인 검증자에서의 딥러닝 기반 이상 탐지를 수행하기 위해, 먼저 모델을 학

습하고 테스트하기 위한 검증자 데이터를 수집했다. 그 후, 검증자 데이터의 차원 수를 줄이기 위해, 차원 축소 기법을 전처리로 사용했다. 전처리의 결과인 저차원 데이터를 모델의 학습, 테스트 데이터로 활용했고, 실험 결과 이러한 방법론이 검증자 이상 탐지에 충분히 효과적임을 알 수 있었다. 또한 실험을 통해 검증자 이상 탐지에서의 적절한 차원 축소 기법과 저차원 데이터의 차원 수를 밝혔다. 향후 본 논문을 발전시켜 블록체인 검증자뿐만 아니라 월렛, 블록체인 RPC 게이트웨이와 같은 블록체인의 다른 구성요소에서도 이상 탐지를 수행하는 연구를 이어 나갈 예정이다.

REFERENCES

- [1] S. Nakamoto., "Bitcoin: A Peer-to-Peer Electronic Cash System," White paper, 2008.
- [2] M. K. Kim., M. H. Kwon., H. J Lee. and S.M. Oh., "Technical Trends of Identity Authentication in Blockchain-based Distributed Environments," REVIEW OF KIISC, vol. 34, no. 1, pp. 45-52, Feb. 2024.
- [3] D.-M. Kye. and S. H. Hur., "A Study on the Possibility of Blockchain Technology Adoption in the Logistics Industry," J. Korean Soc. Intell. Transp. Syst., vol. 21, no. 2, pp. 116-131, Apr., 2022.
- [4] F. Wang, Y. Jiang, R. Zhang, A. Wei, J. Xie, and X. Pang, "A Survey of Deep Anomaly Detection in Multivariate Time Series: Taxonomy, Applications, and Directions," Sensors, vol. 25, no. 1, p. 190, Jan. 2025.
- [5] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," IEEE Access, vol. 9, pp. 120043-120065, Aug. 2021.
- [6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Surveys (CSUR), vol. 41, no. 3, pp. 1-58, Jul. 2009.
- [7] A. Salau, R. Dantu, K. Morozov, S. Badraddoja, and K. Upadhyay, "Making Blockchain Validators Honest," in Proc. 2022 Fourth International Conference on Blockchain Computing and Applications (BCCA), pp. 267-273, San Antonio, USA, Nov. 2022.
- [8] L. Marchesi, M. Marchesi, R. Tonelli, and M. I. Lunesu, "A blockchain architecture for

- industrial applications," Blockchain: Research and Applications, vol. 3, no. 4, p. 100088, Dec. 2022.
- [9] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long Short Term Memory Networks for Anomaly Detection in Time Series," in Proc. 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 89-94 Bruges, Belgium, Apr. 2015.
- [10] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," IEEE Access, vol. 7, pp. 1991-2005, Dec. 2018.
- [11] M. Munir, S. A. Siddiqui, M. A. Chattha, A. Dengel, and S. Ahmed, "FuseAD: Unsupervised Anomaly Detection in Streaming Sensors Data by Fusing Statistical and Deep Learning Models," Sensors, vol. 19, no. 11, p. 2451, May. 2019.
- [12] D. Park, Y. Hoshi, and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based Variational Autoencoder," IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 1544-1551, Jul. 2018.
- [13] H. Kang and P. Kang, "Transformer-Based Multivariate Time Series Anomaly Detection Using Inter-Variable Attention Mechanism," Knowledge-Based Systems, vol. 290, p. 111507, Apr. 2024.
- [14] A. Anuragi, D. S. Sisodia, and R. B. Pachori, "Mitigating the Curse of Dimensionality Using Feature Projection Techniques on Electroencephalography Datasets: An Empirical Review," Artificial Intelligence Review, vol. 57, p. 75, Feb. 2024.
- [15] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," Complex & Intelligent Systems, vol. 8, pp. 2663-2693, Jan. 2022.
- [16] A. Maćkiewicz and W. Ratajczak, "Principal Components Analysis (PCA)," Computers & Geosciences, vol. 19, no. 3, pp. 303-342, Mar. 1993.
- [17] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection," Journal of Open Source Software, vol. 3, no. 29, p. 861, Jul. 2018.
- [18] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.
- [19] M. Balasubramanian and E. L. Schwartz, "The Isomap Algorithm and Topological Stability," Science, vol. 295, no. 5552, p. 7, Jan. 2002.
- [20] D. Grandjean, L. Heimbach, and R. Wattenhofer, "Ethereum Proof-of-Stake Consensus Layer: Participation and Decentralization," in Proc. Financial Cryptography and Data Security International Workshops (FC Workshops), pp. 253-280, Willemstad, Curaçao, Mar. 2024.
- [21] S. Kim, K. Choi, H.-S. Choi, B. Lee, and J.-G. Lee, "Towards a Rigorous Evaluation of Time-Series Anomaly Detection," in Proc. AAAI Conference on Artificial Intelligence, pp. 7194-7201, Jun. 2022.

 저 자 소 개



이 승 재(학생회원)
2025년 한성대학교 IT융합공학부
학사 졸업.
2025년~현재 한성대학교
AI응용학과 석사과정.

<주관심분야: Anomaly detection, Data-Driven AI, LLM>



김 명 선(정회원)
2000년~2002년 LG전자
엑세스네트워크 연구소
주임연구원
2002년~2011년 삼성전자
DMC 연구소 책임연구원
2016년 서울대학교
전기컴퓨터공학부
박사 졸업.

2016년~2019년 삼성전자 SR연구소 수석연구원
2019년~현재 한성대학교 AI응용학과 부교수
<주관심분야: 생성형 AI, 멀티모달 LLM, 데이터
기반 AI, Linux kernel, HW/SW Co-design,
NPU 등>