사물인터넷융복합논문지 Vol. 12, No. 4, pp. 13-18, 2026 https://doi.org/10.20465/KIOTS.2026.12.4.013

소규모 데이터 환경에서의 기계 학습 모델 일반화 성능 비교 연구

이현섭*
백석대학교 컴퓨터공학부 교수

A Comparative Study on the Generalization Performance of Machine Learning Model in Small-Scale Data Environments

Hyun-Seob Lee*
Professor, Division of Computer Engineering, Baekseok University

요 약 최근 사물인터넷(Internet of Things, IoT) 환경의 확산으로 인해 다양한 센서 기반 데이터가 생성되고 있으며, 이를 활용한 지능형 분류 및 예측 모델에 대한 연구가 활발히 진행되고 있다. 그러나 실제 IoT 응용 환경에서는 데이터 수집 비용, 보안 문제, 장비 제약 등의 이유로 대규모 데이터 확보가 어려운 경우가 빈번하게 발생한다. 이러한 소규모 데이터 환경에서는 딥러닝 기반 모델이 요구하는 방대한 학습 데이터를 충족시키기 어렵고, 과적합(overfitting) 문제가 심각하게 나타날 수 있다. 본 연구의 목적은 소규모 IoT 데이터 환경에서 랜덤 포레스트의 일반화 성능과 성능 안정성을 실험적으로 분석하는 데 있다. 이를 위해 Kaggle에서 제공되는 Small IoT Dataset을 활용하여 랜덤 포레스트, 서포트 벡터 머신(support vector machine), 신경망 모델을 대상으로 비교 실험을 수행한다. 특히 단일 학습-검증 분할 방식이 아닌 교차검증(cross-validation) 기반 평가를 적용함으로써, 각 모델의 평균 성능뿐 아니라 성능 분산을 함께 분석하여 일반화 능력을 정량적으로 비교한다. 이러한 분석은 급격하게 변화하는 소규모 사물인터넷 환경에서 기계학습을 적용하는데 기여할 수 있을 것으로 기대한다.

주제어 : 사물인터넷, 소규모 데이터셋, 기계학습, 데이터 부족, 과적합 방지

Abstract The proliferation of the Internet of Things (IoT) environment has led to the generation of diverse sensor-based data, and research on intelligent classification and prediction models utilizing this data is actively underway. However, in real IoT application environments, acquiring large-scale data is often difficult due to factors such as data collection costs, security concerns, and equipment limitations. In these small-scale data environments, it is challenging to meet the vast training data requirements of deep learning-based models, and overfitting issues can become severe. The objective of this study is to experimentally analyze the generalization performance and stability of Random Forest in small-scale IoT data environments. To achieve this, comparative experiments are conducted using the Small IoT Dataset provided by Kaggle, evaluating Random Forest, Support Vector Machine (SVM), and neural network models. Specifically, by applying cross-validation-based evaluation instead of a single training-validation split, we quantitatively compare the generalization capabilities of each model by analyzing not only their average performance but also their performance variance. This analysis is expected to contribute to the application of machine learning in rapidly changing small-scale IoT environments.

Key Words : IoT, Small Dataset, Machine Learning, Data Poverty, Overfitting Prevention

This paper was supported by 2026 Baekseok University Research Fund
*교신저자 : 이현섭(hyunseob@bu.ac.kr)
접수일 2026년 02월 17일 수정일 2026년 06월 06일 심사완료일 2026년 06월 12일

1. 서론

최근 사물인터넷(Internet of Things, IoT) 환경의 확산으로 인해 다양한 센서 기반 데이터가 생성되고 있으며, 이를 활용한 지능형 분류 및 예측 모델에 대한 연구가 활발히 진행되고 있다[1-4]. 그러나 실제 IoT 응용 환경에서는 데이터 수집 비용, 보안 문제, 장비 제약 등의 이유로 대규모 데이터 확보가 어려운 경우가 빈번하게 발생한다. 이러한 소규모 데이터 환경에서는 딥러닝 기반 모델이 요구하는 방대한 학습 데이터를 충족시키기 어렵고, 과적합(overfitting)[5, 6] 문제가 심각하게 나타날 수 있다. 그러나, 기존 머신러닝 및 딥러닝 연구들은 주로 대규모 데이터셋을 가정하고 모델의 성능을 평가해 왔다. 특히 신경망(neural network)[7, 8] 기반 모델은 충분한 학습 데이터가 제공될 경우 우수한 성능을 보이지만, 데이터 수가 제한적인 환경에서는 성능 변동성이 커지고 일반화 성능이 저하되는 경향이 있다. 반면, 앙상블 학습 방법 중 하나인 랜덤 포레스트(random forest)[9, 10]는 배깅(bagging)과 무작위 특성 선택을 통해 분산을 감소시키는 구조적 특성으로 인해 상대적으로 소규모 데이터 환경에서도 안정적인 성능을 보이는 것으로 알려져 있다. 그러나 이러한 장점에도 불구하고, 소규모 IoT 데이터 환경에서 랜덤 포레스트의 일반화 성능을 다른 대표적 분류 모델과 체계적으로 비교·분석한 연구는 충분하지 않다. 본 연구의 목적은 소규모 IoT 데이터 환경에서 랜덤 포레스트의 일반화 성능과 성능 안정성을 실험적으로 분석하는 데 있다. 이를 위해 Kaggle에서 제공되는 Small IoT Dataset을 활용하여 랜덤 포레스트, 서포트 벡터 머신(support vector machine)[11, 12], 신경망 모델을 대상으로 비교 실험을 수행한다. 특히 단일 학습–검증 분할 방식이 아닌 교차검증(cross-validation) 기반 평가를 적용함으로써, 각 모델의 평균 성능뿐 아니라 성능 분산을 함께 분석하여 일반화 능력을 정량적으로 비교한다. 본 연구에서는 Stratified K-Fold 교차검증을 사용하여 클래스 분포를 유지한 상태에서 모델 성능을 평가한다. 평가 지표로는 분류 정확도를 사용하며, 각 모델에 대해 교차검증 결과의 평균과 표준편차를 산출한다. 또한 박스플롯, 평균±표준편차 바 차트, 학습 곡선(learning curve) 등 다양한 시각화 방법을 통해 모델의 성능 안정성과 데이터 크기에 따른 일반화 특성을 분석한다.

2. 배경

최근 머신러닝 및 딥러닝 기술의 발전으로 다양한 산업 분야에서 데이터 기반 의사결정 시스템이 활용되고 있다. 특히 사물인터넷(IoT) 환경에서는 센서를 통해 수집된 데이터를 기반으로 이상 탐지, 상태 진단, 사용자 행동 예측 등 다양한 응용이 이루어지고 있다. 그러나 실제 IoT 환경에서는 데이터 수집 비용, 저장 공간의 한계, 개인정보 보호 문제 등으로 인해 대규모 학습 데이터를 확보하기 어려운 경우가 많다. 이러한 소규모 데이터 환경에서는 모델의 과적합 가능성이 증가하며, 모델의 일반화 성능이 중요한 평가 요소가 된다. 랜덤 포레스트는 다수의 결정트리를 배깅 방식으로 결합하는 앙상블 학습 기법으로, 데이터의 무작위 샘플링과 특성 무작위 선택을 통해 분산을 감소시키는 특징을 가진다. 이러한 구조적 특성은 개별 결정트리의 높은 분산을 완화하고, 비교적 적은 데이터 환경에서도 안정적인 예측 성능을 유지하도록 돕는다. 반면, 서포트 벡터 머신은 마진 최대화 원리를 기반으로 하여 소규모 데이터에서도 비교적 우수한 성능을 보이는 것으로 알려져 있으며, 신경망은 충분한 데이터가 확보될 경우 강력한 표현력을 발휘하지만 데이터가 제한적인 경우 과적합 문제가 발생할 수 있다. 따라서 소규모 데이터 환경에서 각 모델이 보이는 일반화 성능의 특성을 비교·분석하는 것은 실용적 및 이론적 측면에서 모두 중요하다.

3. 관련 연구

3.1 랜덤 포레스트와 앙상블

랜덤 포레스트(random forest)는 앙상블 학습 기법으로, 다수의 결정트리를 부트스트랩 샘플링을 통해 생성하고 각 노드 분할 시 무작위로 선택된 특성 부분집합을 사용함으로써 모델 간 상관성을 감소시키는 구조를 갖는다. 이러한 배깅 기반 구조는 개별 결정트리의 높은 분산(high variance) 문제를 완화하며, 전체 모델의 일반화 오차를 감소시키는 효과를 가진다. 이론적으로 랜덤 포레스트의 일반화 오차는 개별 트리의 강도(strength)와 트리 간 상관성(correlation)에 의해 결정되며, 트리 간 상관성이 낮을수록 앙상블의 성능은 향상된다. 이러한 특성은 데이터 수가 제한된 환경에서도 과적합 위험을 비교적 효과적으로 제어할 수 있도록 한다. 실제로 다

양한 응용 연구에서 랜덤 포레스트는 노이즈가 존재하거나 표본 수가 충분하지 않은 데이터 환경에서도 안정적인 예측 성능을 보이는 것으로 보고되었다. 특히 의료 진단, 생물정보학, 금융 리스크 분석과 같이 표본 수는 제한적이지만 변수 차원이 상대적으로 높은 문제에서 모델 복잡도 조정 없이도 상대적으로 안정적이고 우수한 성능을 보이는 대표적 기법으로 활용되어 왔다.

3.2 서포트 벡터 머신과 구조적 위험 최소화

서포트 벡터 머신(support vector machine, SVM)은 통계적 학습 이론에 기반한 분류 기법으로, 구조적 위험 최소화(structural risk minimization)를 수행하는 특징이 있다. SVM은 마진(margin)을 최대화하는 초평면을 찾음으로써 일반화 성능을 향상시키며, 특히 표본 수가 적은 경우에도 비교적 안정적인 성능을 보인다. 또한 커널 트릭(kernel trick)을 활용하여 비선형 결정 경계를 효과적으로 모델링할 수 있어, 복잡한 데이터 구조를 가진 소규모 데이터셋에서도 높은 분류 정확도를 달성할 수 있다. 실제로 의료 영상 분석, 유전자 데이터 분류, 텍스트 분류와 같은 소규모 데이터 응용 분야에서 SVM은 오랫동안 표준적인 기법으로 활용되어 왔다. 그러나 SVM은 하이퍼파라미터(C, gamma 등)에 민감하며, 데이터 규모가 매우 작거나 노이즈가 많은 경우 결정 경계가 불안정해질 수 있다는 한계도 존재한다. 따라서 교차검증 기반의 성능 안정성 분석이 필요하다.

3.3 신경망 모델과 소규모 데이터의 한계

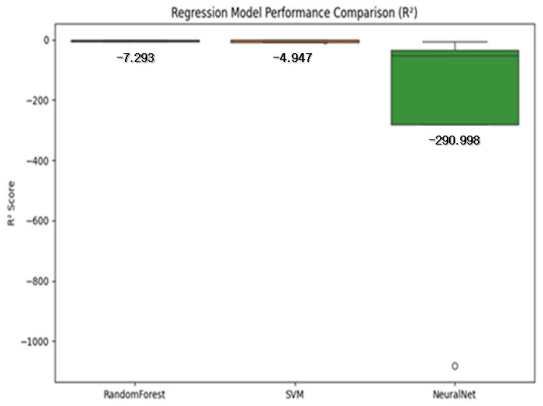
신경망(neural network)은 다층 구조를 통해 복잡한 비선형 관계를 모델링할 수 있으며, 최근 딥러닝의 발전과 함께 다양한 분야에서 최첨단 성능을 달성하고 있다. 그러나 신경망은 일반적으로 많은 수의 파라미터를 포함하고 있어 충분한 학습 데이터가 확보되지 않을 경우 과적합 문제가 발생하기 쉽다. 소규모 데이터 환경에서는 모델의 표현력에 비해 학습 샘플이 부족하여 학습 오차는 낮지만 검증 오차가 증가하는 현상이 나타날 수 있다. 이를 완화하기 위해 정규화(regularization), 드롭아웃(dropout), 조기 종료(early stopping) 등의 기법이 제안되어 왔으나, 이러한 방법들은 추가적인 하이퍼파라미터 조정과 반복 실험을 필요로 한다. 최근 일부 연구에서는 데이터 증강(data augmentation)이나 전이학습(transfer learning)을 통해 소규모 데이터 문제를 해결하려는 시도가 이루어지고 있으나, 이는 주로 이미지나 대규모 사전학습 모델이 존재하는 영역에 국한되는 경향이 있다. 따라서, 일반적인 구조화된 IoT 데이터 환경에서는 이러한 접근이 제한적일 수 있다.

3.4 소규모 데이터 환경에서 일반화 성능 비교

기존 연구들은 주로 대규모 벤치마크 데이터셋에서 모델의 평균 정확도를 비교하는 데 초점을 두어 왔다. 그러나 소규모 데이터 환경에서는 단일 학습–검증 분할 방식이 성능을 과대 혹은 과소 추정할 가능성이 크며, 모델의 평균 성능뿐 아니라 성능 분산 또한 중요한 평가 지표가 된다. 일부 연구에서는 교차검증을 활용하여 모델의 일반화 성능을 평가하였으나, 소규모 IoT 데이터 환경에서 랜덤 포레스트, SVM, 신경망을 동일 조건 하에 비교하고 성능 안정성까지 체계적으로 분석한 연구는 부족하다. 특히 박스플롯, 학습곡선(learning curve), 평균 및 표준편차 비교를 통해 모델의 안정성과 데이터 크기 민감도를 종합적으로 분석한 연구는 충분히 이루어지지 않았다. 따라서 본 연구는 소규모 IoT 데이터셋을 활용하여 세 가지 대표적 분류 모델의 교차검증 기반 일반화 성능과 성능 변동성을 정량적으로 비교함으로써, 소규모 데이터 환경에서 모델 선택 기준에 대한 정량적인 근거를 분석한다.

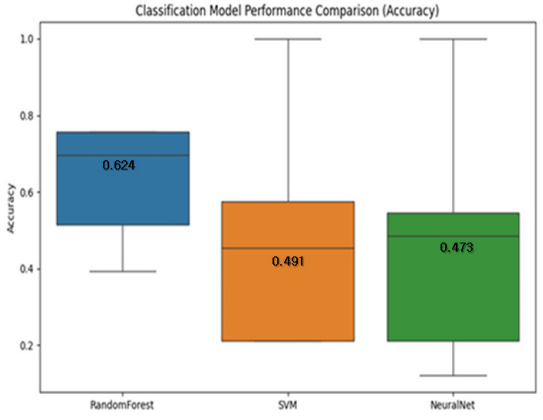
4. 소규모 환경에서 일반화 성능 분석

본 연구는 소규모 IoT 데이터 환경에서 분류 모델의 일반화 성능과 성능 안정성을 체계적으로 비교·분석하는 것을 목표로 한다. 기존 연구들이 주로 대규모 데이터셋에서 평균 정확도를 중심으로 모델을 비교해 온 것과 달리, 본 연구는 데이터 수가 제한된 환경에서 모델의 성능 분산과 안정성에 초점을 둔다. 이를 위해 랜덤 포레스트, 서포트 벡터 머신(SVM), 신경망 세 가지 대표적 분류 모델을 동일한 조건에서 학습시키고, Stratified K-Fold 교차검증을 기반으로 평균 성능뿐 아니라 표준편차, 학습곡선, 데이터 크기 변화에 따른 성능 추이를 종합적으로 분석한다. 성능 분석환경은 AMD Ryzen Threadripper PRO 5975WX 32-Cores, 3600Mhz, 512GB RAM, RTX4090x4를 탑재한 윈도우 OS 환경의 워크스테이션이다. 분석에 사용된 데이터는 Kaggle에 공개된 소규모 IoT 환경에서 수집된 데이터[15]다.



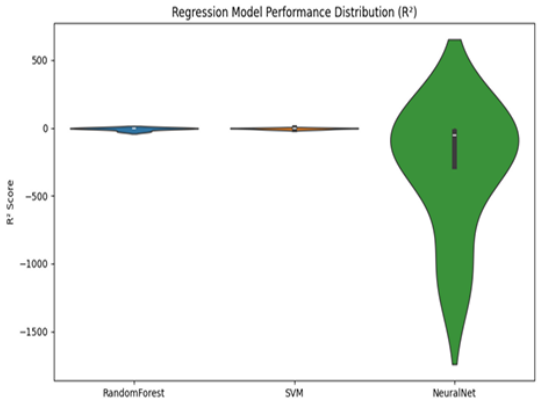
[Fig. 1] Regression Model Performance Comparison

Fig. 1은 랜덤 포레스트, SVM 신경망에서 소규모 데이터를 이용한 Regression의 성능을 보여주고 있다. y축은 각 모델별 R^2 점수다. 이 점수는 변동 폭이 작고, 밀집될수록 안정적인 성능을 의미한다. 그림의 결과와 같이 SVM이 평균 -4.947로 가장 양호한 성능을 보였으며, 랜덤 포레스트가 -7.293를 기록하였다. 반면, 신경망은 평균 -290.998로 매우 낮고 불안정한 수치를 기록여 소규모 데이터에서 모델이 수렴하지 못하거나 과적합되었음을 의미한다.



[Fig. 2] Classification Model Performance Comparison

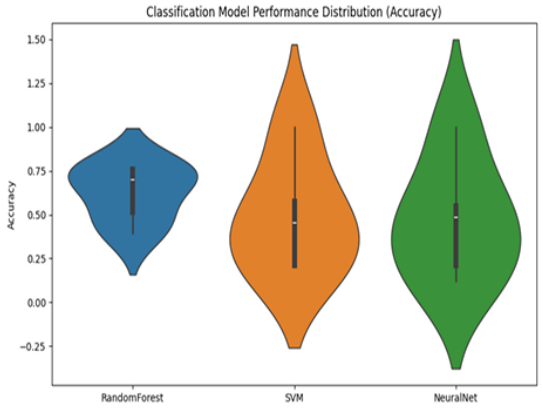
Fig. 2는 Classification의 성능을 박스플롯 그래프로 보여주고 있다. 그림에서 y축은 각 모델별 정확도를 의미한다. 정확도 측면에서 랜덤 포레스트의 범위가 가장 좁고 평균값은 0.624로 가장 우수하고 안정적인 성능을 보였다. SVM과 신경망은 정확도가 1.0에 도달하는 경우도 있지만, 각각 0.491과 0.473의 평균 정확도를 기록하여 상대적으로 낮은 평균 정확도를 기록했다. 또한, 최고점과 최저점 간의 편차가 커서 정확도의 기복이 있음을 확인할 수 있었다.



[Fig. 3] Data Density on Regression

Fig. 3는 앞선 박스 플롯을 이용한 Regression 모델 성능 분석에 데이터의 밀도를 시각화한 바이올린 플롯 그래프를 보여주고 있다. 이 그래프는 데이터가 집중된 구간을 시각적으로 보여준다. R^2 점수는 1에 가까울수록 완벽한 예측을 의미하며, 음수 값은 모델이 데이터의 평균값보다 예측을 못하고 있음을 의미한다. 그림의 결과에서 랜덤 포레스트의 표준편차는 9.778이고 최댓값과 최솟값은 각각 -0.95, -26.636이다. 또한 0 근처에서 바이올린의 배가 형성되어 있어 대부분의 실험에서 상대적으로 일정한 성능을 유지하고 있음을 보여준다. SVM의 표준편차는 4.511로 세 모델 중 평균적으로 가장 좋은 성능을 보였다. 바이올린의 폭 또한 가장 좁고 0에 가깝게 뭉쳐 있어 변동성이 적고 가장 안정적인 예측을 수행했음을 알 수 있다. 마지막으로, 신경망은 최솟값이 -1081.765로 극단적인 결과를 보여주고 있다. 바이올린이 아래 음수 방향으로 길게 늘어진 형태를 보여주고 있다. 이러한 결과는 소규모 데이터 환경에서 신경망이 특정 데이터(Fold 5)에 대해 완전히 잘못된 학습을 하여 성능이 심하게 저하되는 구간이 존재함을 보여주고 있다.

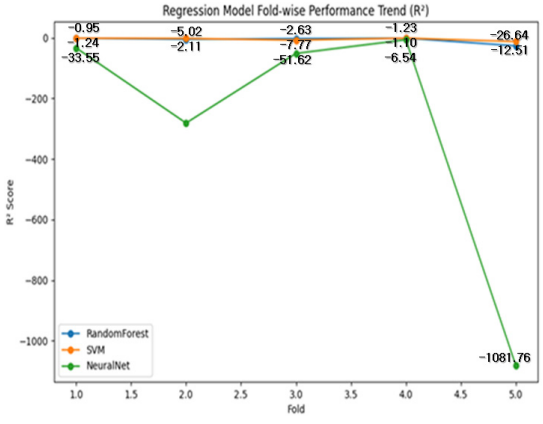
Fig. 4는 Classification 모델의 성능분석 결과에 데이터 밀도를 시각화 한 결과다. Classification 실험에서 정확도는 0에서 1사이의 값을 가지며 1에 가까울수록 우수한 성능을 의미한다. 그림에서 랜덤 포레스트의 바이올린의 몸통이 0.6에서 0.7 구간에 대다수의 데이터가 형성되어 있고, 최솟값이 0.394로 다른 모델보다 높다. 이러한 결과는 데이터가 적은 환경에서도 최소한의 성능을 보장하는 모델임을 의미한다. 반면 SVM과 신경망 모



[Fig. 4] Data Density on Classification

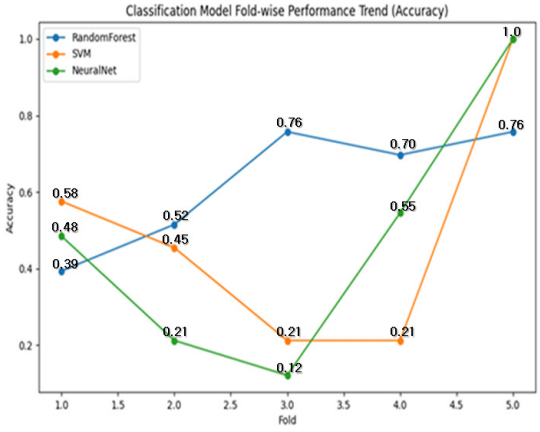
델은 바이올린이 상하로 길게 뻗어 있다. 또한, SVM은 성능이 좋을 때 1, 나쁠 때 0.21의 정확도를 보여주고 있어 데이터에 따라 성능 기복이 심한 결과를 보여주었고, 신경망은 최저점 부근에서도 데이터 밀도가 감지되어, 소규모 임베디드 데이터 환경에서는 다른 모델보다 성능이 저하될 위험이 크다.

성능결과와 바이올린 플롯 분석을 통해 신경망은 데이터 부족 시 성능이 극단적으로 낮아지는 긴 꼬리 현상을 보이는 반면, 랜덤 포레스트는 적은 데이터 환경에서도 특정 구간에 성능이 밀집되고 안정적인 형태를 유지하는 것을 알수 있다. 이러한 실험을 통해 소규모 데이터 환경에서는 신경망 보다 전통적 머신 러닝이 더 적합할 수 있다는 것을 예상할 수 있다.



[Fig. 5] Regression R2 Score Trend

Fig. 5는 Regression 모델의 성능 추이를 보여주고 있다. 그림에서 x축은 5-Fold 교차 검증의 각 단계를 나타내며, 단계별로 모델의 R^2 성능이 어떻게 변하는지 시계열로 보여주고 있다. 랜덤 포레시트는 Fold1에서 -0.95로 시작해서 Fold5에서 -26.63으로 하락하지만, 전반적으로 0에 가까운 선을 유지하며 급격한 변동 없이 일정한 성능을 보여주었다. SVM은 -33.55로 시작하여 Fold4까지는 -1.10로 안정적인 성능을 보여주었고, Fold5에서 -12.51로 소폭 하락하였으나 세 모델 중 가장 변동 폭이 적고 안정적인 성능을 보여주었다. 반면 신경망은 -33.55로 시작해서 Fold5에서는 -1081.76으로 급격히 하락하였다. 이러한 결과는 소규모 데이터 환경에서 신경망이 특정 패턴에 따라 성능이 크게 저하될 수 있음을 의미한다.



[Fig. 6] Classification Accuracy Trend

Fig. 6은 Classification 모델의 성능 추이를 보여주고 있다. 그림의 결과와 같이 랜덤 포레스트는 각 Fold 별로 0.39, 0.52, 0.76, 0.70, 0.76의 정확도를 보여주었고 다른 모델의 성능이 낮을 때에도 0.4 이상의 견고한 성능을 보여주었다. SVM은 0.58에서 시작하였으나, Fold3, 4에서 0.21까지 저하되었다가 최종 1.0까지 V자 형태의 성능 변화를 보여주었다. 마지막으로 신경망은 0.48로 시작하여 0.12까지 저하되었다가 1.0으로 가파른 상승결과를 보여주었다. 이러한 결과는 데이터량이 적어서 각 폴드에 과적합 된 것으로 판단된다. 실험을 통해 신경망이 소규모 데이터 환경에서는 성능이 극단적으로 변하는 불안정을 보였고, SVM은 일부 데이터 패턴에 민감한 결과를 보였다. 반면 랜덤포레스트는 상대적으로 큰 변화 없는 성능을 보여주었다.

5. 결론

본 연구는 소규모 IoT 데이터 환경에서 랜덤 포레스트, SVM, 신경망 모델의 일반화 성능을 교차검증 기반으로 비교하였다. 실험 결과, 신경망은 데이터가 제한된 환경에서 성능 저하와 높은 변동성을 보였으며, 학습곡선 분석에서도 과적합 경향이 확인되었다. 반면, 랜덤 포레스트는 평균 성능과 성능 안정성 측면에서 비교적 우수한 결과를 보였으며, 데이터 크기 감소 구간에서도 성능 저하 폭이 작게 나타났다. 이는 앙상블 기반 구조가 분산을 감소시키고 과적합을 완화한 결과로 해석된다. 본 연구는 소규모 데이터 환경에서는 복잡한 고용량 모델보다 전통적인 기계학습 모델이 더 안정적인 일반화 성능을 제공할 수 있음을 증명하였다. 향후에는 다양한 소규모 데이터 환경에 적합한 기계학습을 연구할 예정이다.

REFERENCES

[1] M.Mohammadi, A.Al-Fuqaha, S.Sorour and M.Guizani, "Deep learning for IoT big data and streaming analytics: A survey," IEEE Communications Surveys & Tutorials, Vol.20, No.4, pp.2923-2960, 2018.

[2] M.S.Mahdavinejad, M.Rezvan, M.Barekatain, P.Adibi, P.Barnaghi and A.P.Sheth, "Machine learning for internet of things data analysis: A survey," Digital Communications and Networks, Vol.4, No.3, pp.161-175, 2018.

[3] M.A.Al-Garadi, A.Mohamed, A.K.Al-Ali, X.Du, I.Ali and M.Guizani, "A survey of machine and deep learning methods for internet of things (IoT) security," IEEE Communications Surveys & Tutorials, Vol.22, No.3, pp.1646-1685, 2020.

[4] M.Shafiq, Z.Tian, A.K.Bashir, X.Du and M.Guizani, "CorrAUC: A malicious bot-IoT traffic detection method in IoT network using machine learning techniques," IEEE Internet of Things Journal, Vol.8, No.5, pp.3242-3254, 2021.

[5] C.F.G.Santos and J.P.Papa, "Avoiding overfitting: a survey on regularization methods for convolutional neural networks," ACM Computing Surveys, Vol.54, No.10, pp.1-25, 2022.

[6] Z.S.Kadhim, H.S.Abdullah and K.I.Ghathwan, "Automatically avoiding overfitting in deep neural networks by using hyper-parameters optimization methods," International Journal of Online and Biomedical Engineering, Vol.19, No.4, pp.4-18, 2023.

[7] L.Alzubaidi, J.Zhang, A.J.Humaidi, A.Al-Dujaili, Y.Duan, O.Al-Shamma, J.Santamaría, M.A.Fadhel, M.Al-Amidie and L.Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data, Vol.8, No.1, pp.53, 2021.

[8] Z.Li, F.Liu, W.Yang, S.Peng and J.Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE Transactions on Neural Networks and Learning Systems, Vol.33, No.12, pp.6999-7019, 2022.

[9] N.M.Abdulkareem and A.M.Abdulazeez, "Machine learning classification based on random forest algorithm: a review," International Journal of Science and Business, Vol.5, No.2, pp.128-142, 2021.

[10] M.Schonlau and R.Y.Zou, "The random forest algorithm for statistical learning," Stata Journal, Vol.20, No.1, pp.3-29, 2020.

[11] J.Cervantes, F.Garcia-Lamont, L.Rodríguez-Mazahua and A.Lopez, "A comprehensive survey on support vector machine classification: applications, challenges and trends," Neurocomputing, Vol.408, pp.189-215, 2020.

[12] J.Wang, F.He and S.Sun, "Construction of a new smooth support vector machine model and its application in heart disease diagnosis," PLOS ONE, Vol.18, No.2, pp.e0280804, 2023.

[13] V.W.Lumumba, D.Kiprotich, M.L.Mpaine, N.G.Makena and M.D.Kavita, "Comparative analysis of cross-validation techniques: LOOCV, K-folds cross-validation, and repeated K-folds cross-validation in machine learning models," American Journal of Theoretical and Applied Statistics, Vol.13, No.5, pp.127-137, 2024.

[14] S.Prusty, S.Patnaik and S.K.Dash, "SKCV: stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," Frontiers in Nanotechnology, Vol.4, pp.972421, 2022.

[15] https://www.kaggle.com/datasets/mertkalkanci/small-iot-dataset

이 현 섭(Hyun-Seob Lee) [종신회원]

- 2013년 2월 : 한양대학교 컴퓨터공학과 (공학 박사)
- 2012년 3월 ~ 2021년 2월 : 삼성전자 책임연구원
- 2021년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 조교수

〈관심분야〉
인공지능, 저장시스템, 임베디드 시스템