시간대를 고려한 SHAP 기반의 신용카드 이상 거래 탐지

양소연 [1], 임유진 [2]
[1] 숙명여자대학교 빅데이터분석융합학(협동과정)
[2] 숙명여자대학교 인공지능공학부
syyang@sookmyung.ac.kr, yujin91@ sookmyung.ac.kr

Credit Card Fraud Detection Based on SHAP Considering Time Sequences

Soyeon yang[1], Yujin Lim[2]
[1]Dept. of Big Data Analysis Convergence, Sookmyung Women's University
[2]Division of Artificial Intelligence Engineering, Sookmyung Women's University

요 약

신용카드 부정 사용은 고객 및 기업의 신용과 재산에 막대한 손실을 미치고 있다. 이에 따라 금융사들은 이상금융거래탐지시스템을 도입하였으나 이상 거래 발생 여부를 지속적으로 모니터링 하고 있기 때문에 시스템 유지에 많은 비용이 따른다. 따라서 본 논문에서는 컴퓨팅 리소스를 절약함과 동시에 성능 개선 효과를 보인 신용카드 이상 거래 탐지 알고리즘을 제안한다. CTGAN 을 활용하여 정상 거래와 이상 거래의 비율을 일부 완화하였고 XAI 기법인 SHAP 를 활용하여 유의미한 속성값을 선택하였다. 이것을 기반으로 LSTM Autoencoder 를 사용하여 이상데이터를 탐지하였다. 그 결과 전통적인 비지도 학습 기법에 비해 제안 알고리즘이 우수한 성능을 보였음을 확인하였다.

1. 서론

신용카드 부정 사용이란 신용카드 위변조, 타인의 정보 유출 및 이용 등 전반적인 카드 업무 중에서 발생하는 비정상적인 사용 행태를 말한다[1]. 이런 부정 사용 행태는 고객뿐만 아니라 기업에도 큰 손해를 야기할 수 있기에 대다수의 은행에서는 이상거래탐지시스템(FDS, Fraud Detection System)을 사용하여 비정상적인 사용 행태를 탐지하고 있다. 하지만 정상 거래인 것처럼 교묘하게 거래 승인을 취하거나 간편 거래 등의 보안상의 취약점을 노린 거래가 있다면 사전에 조치를 취하는 것이 어렵다[1]. 이상치 탐지 알고리즘을 제안하기 위해 본 논문에서는 이상 거래와 정상 거래의 데이터 불균형(data imbalance) 문제를 고려하였고 효과적인 변수 선택(feature selection)을 통해 탐지 모델의 훈련 시간 및 비용을 절약하고자 하였다.

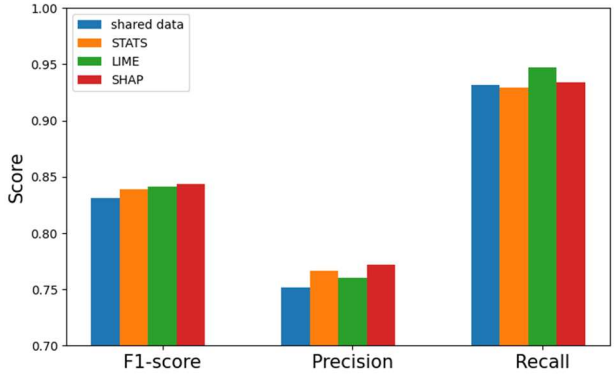
대다수의 FDS 모델은 이상치 비율이 정상치 비율보다 압도적으로 적기 때문에 데이터 불균형 문제를 겪고 있다. 이러한 문제는 모델의 학습 성과를 저하시키며 해결 방법으로는 언더샘플링(under-sampling) 기법과 오버샘플링(over-sampling) 기법이 있다. 언더샘플링 기법은 정상치의 비율을 줄임으로써 이상치 비율과 동일하게 만드는 기법이다. 이 방법은 데이터 불균형 문제를 해결할 수는 있지만 전체적인 데이터 수를 감소시키므로 데이터의 특성을 제거한다는 단점이 있다. 이와 달리 오버샘플링 기법은 이상치 비율을 늘림으로써 정상치 비율과 동일하게 만드는 방법으로 SMOTE(Synthetic Minority Over-sampling Technique)와 GAN(Generative Adversarial Network)등의 기법이 사용되고 있다. 모델의 성능을 개선하고 비용을 절약하기 위한 변수 선택 방법으로 XAI(eXplainable Artificial Intelligence)를 사용하였다. XAI 의 목적은 인간에게 AI 의 행동을 이해하기 쉽게 설명해주는 것이며 대표적인 기법으로는 LIME 과 SHAP 가 있다. 해당 논문에서는 날이 갈수록 고도화되는 사기 수법에 대응하기 위해 시계열 모델인 LSTM(Long Short-Term Memory) 기법을 사용하여 이상 거래를 탐지하였다. 제안 알고리즘의 성능을 측정하기 위하여 F1-score, precision, 그리고 recall 값을 계산하여 비교하였다. 특히 precision 과 recall 은 상충(trade-off) 관계이므로 이 둘의 조화 평균인 F1-score 을 중심으로 살펴보고자 한다.

2. 제안 알고리즘

신용카드 이상 거래 탐지를 위해 본 논문에서 사용한 데이터셋[2]은 2013 년 유럽에서 사용된 신용카드

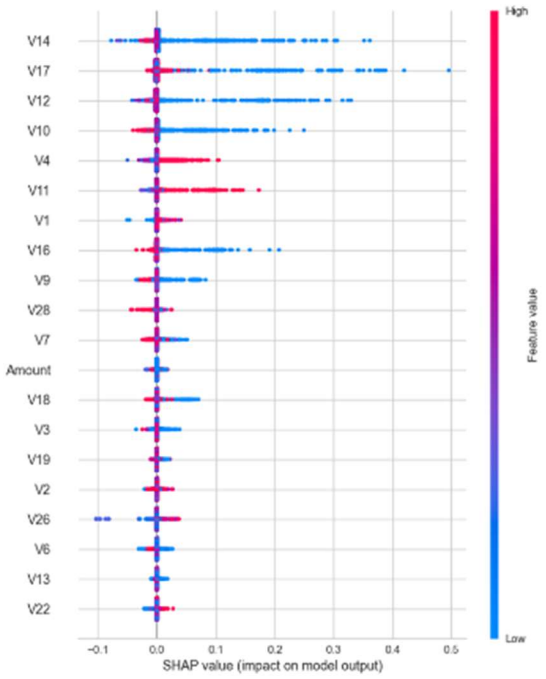
거래 데이터로서 ULB(Université Libre de Bruxelles)에서 수집 및 분석하였다. 해당 데이터 셋에서의 총 거래 건 수는 약 28 만개이며 이중에서 이상 거래는 492 건으로 전체의 약 0.172%를 차지한다. 해당 데이터셋은 이진 분류 기법을 사용하기 때문에 이상 거래와 정상 거래의 클래스 비율이 모델링 성능에 큰 영향을 미친다. 오버샘플링 기법 중 CGAN(conditional GAN)을 활용한 기법이 기존의 ROS(random oversampling)와 SMOTE 보다 뛰어나다[3]는 연구 결과에 기반하여 CTGAN(conditional tabular GAN)으로 오버샘플링 하였다. 다만 이상치와 정상치의 비율을 동일하게 만드는 것은 현실성이 다소 떨어진다고 판단하였기 때문에 전체 비율의 약 0.172%였던 이상치를 각각 0.3%와 0.7%로 설정하여 클래스 불균형을 일부 해소하였다.

효과적인 변수 선택을 위해 피어슨 상관 계수, 카이제곱 검정, 차이분석 등의 통계적 기법(STATS)과 XAI 기법인 LIME 과 SHAP 로 선택한 속성값을 랜덤 포레스트(Random Forest)를 사용하여 비교 및 분석하였다. 랜덤포레스트를 사용한 이유는 K-최근접 이웃, 결정 트리, 서포트 벡터 머신 등 약 10 개의 분류 모델과 비교하였을 때 가장 정확도가 높은 모델이었기 때문이다. 그 결과 SHAP 를 사용하여 속성값을 추출하였을 때의 F1-score 가 약 0.839 로 가장 우수하였음을 확인하였다(그림 1). LIME 과 SHAP 를 활용하여 선택한 속성 추출 값의 성능이 비슷하다는 연구 결과[4]가 있으나 LIME 은 국소 회귀의 적합성만을 고려하는 반면 SHAP 는 전체 데이터 샘플을 활용하여 설명을 제공하기 때문에 해당 연구에서는 전체 데이터 셋을 고려한 SHAP 를 사용하였다.

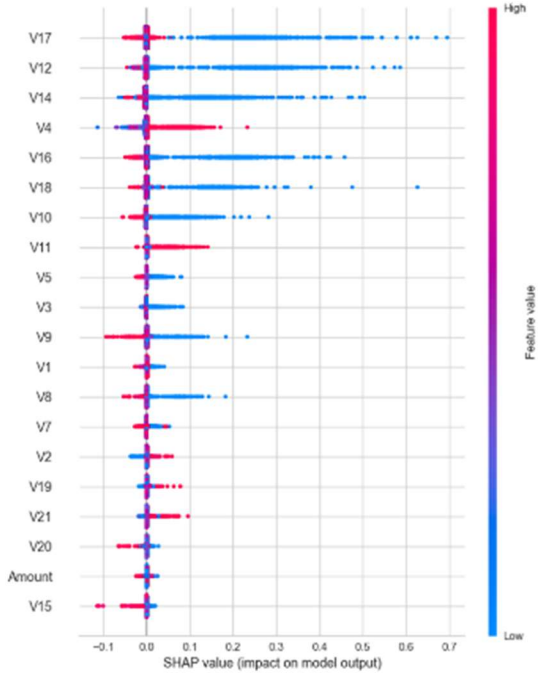


(그림 1) Feature selection 기법에 따른 성능 비교

(그림 2)는 이상치 데이터가 전체 데이터의 약 0.17%를 차지하는 경우(a)와 약 0.7%를 차지하는 경우(b)에서 주요 속성값을 추출하기 위해 SHAP 를 활용한 결과다. CTGAN 을 활용하여 이상 거래 비율이 약 0.7%가 되도록 오버샘플링한 데이터에서도 원본 데이터와 비슷한 양상으로 변수의 중요도가 나타났다.



(a) 이상치가 전체 데이터의 약 0.17%인 데이터셋



(b) 이상치가 전체 데이터의 약 0.7%인 데이터셋

(그림 2) SHAP 를 활용하여 추출한 주요 속성값

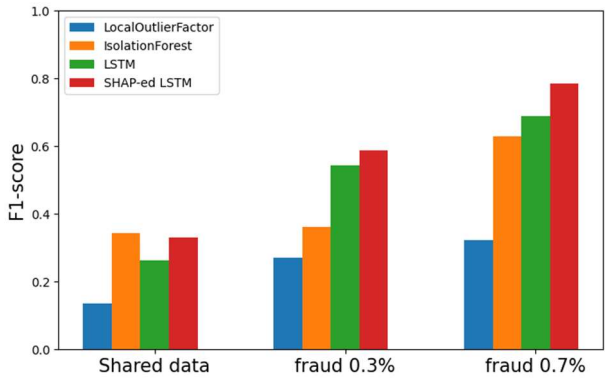
본 논문에서는 데이터 불균형 문제를 해결하기 위해 CTGAN 을 활용하여 오버샘플링하였고 주요 변수값을 추출하기 위하여 SHAP 를 사용하였다. 해당 결과값을 바탕으로 원본데이터의 2/3 개의 속성값만으로 실험을 진행하였으며 시간에 가중치를 주어 변화하는 사기 패턴을 반영하고 이것에 대응하기 위하여 LSTM Autoencoder 모델을 사용하였다.

LSTM 은 Long Short-Term Memory 로서 비교적 과거의 데이터까지 고려하여 예측하는 시계열 모델이고

제안 알고리즘은 이것에 오토인코더(Autoencoder)를 적용하여 학습시키는 모델이다. 오토인코더는 인코더(encoder)와 디코더(decoder)로 구성된 비지도학습 모델로서 인코더에서는 입력 값을 압축하여 변환시키고 디코더에서는 변환된 입력 값을 재생성 시킨다.

3. 실험 및 성능 비교 분석

제안 알고리즘의 우수성을 확인하기 위하여 Local Outlier Factor 와 Isolation Forest 같은 전통적인 비지도 학습 기법과 변수 선택을 하지 않은 LSTM 모델과도 비교 실험을 진행하였다. Local Outlier Factor 는 주어진 데이터의 국소 밀도 편차를 고려하여 이웃한 데이터보다 상당히 낮은 밀도를 가진 표본을 특이치로 간주하는 모델이며 Isolation Forest 는 임의로 데이터의 특징을 선택하고 이것의 최대값과 최소값 사이의 분할값을 랜덤하게 선택하여 각 관측치를 분리시키는 밀도 기반의 트리 모델이다.



(그림 3) 기존 비지도 학습 모델과 제안 알고리즘의 비교

실험에 사용한 데이터셋은 이상치 비율이 약 0.17%인 원본데이터(shared data)와 약 0.3%인 fraud 0.3%, 그리고 0.7%인 fraud 0.7% 데이터이다. 원본데이터로 비지도학습을 하였을 땐 Isolation Forest 와 제안 알고리즘이 비슷한 성능을 보였으나 이상치 비율이 높아질수록 제안 알고리즘인 SHAP-ed LSTM 의 성능이 여타 비지도학습 알고리즘보다 우수하다는 것을 확인하였다.

(그림 4)에서는 시간에 따라 진화하는 범죄 유형에 대응하기 위하여 제안기법에서 LSTM 모델의 시퀀스를 고려하였다. 신용카드 이상 거래 비율이 전체 거래건의 0.3%이하일 때엔 약 50 개의 데이터를 보는 것이 소폭 상승한 결과값을 보였으나 이상치 비율이 0.7% 이상일 때에는 최근 데이터 30 개만 보는 것이 이상치 탐지에 효과적이었다. 이러한 실험 결과로 미루어 보아 이상치 비율이 낮으면 데이터의 전반적인 흐름을 보아야 하지만 이상치 비율이 높으면 최근 데이터만 보는 것이 이상금융거래탐지에 유리함을 알 수 있었다.

(그림 4) 최근 데이터 반영 개수에 따른 모델 성능 비교

4. 결론

본 연구에서는 이상금융거래탐지시스템(FDS)의 컴퓨팅 리소스를 절약하기 위한 방법으로 XAI 기법 중 하나인 SHAP 를 활용하였다. 데이터의 전체 속성값을 모두 사용하는 대신 SHAP 기법으로 추출한 데이터의 속성값만으로도 신용카드 부정 사용 거래를 탐지할 수 있는지를 Local Outlier Factor 와 Isolation Forest 를 사용하여 비교하였다. 비교 실험을 진행하기에 앞서 이상치 탐지 문제에서 겪고 있는 데이터 불균형 문제를 완화하기 위해 CTGAN 을 사용하였다. 그리고 날이 갈수록 점차 교묘해지는 비정상적인 사용 행태를 탐지하기 위하여 LSTM 모델을 사용함으로써 시간의 흐름을 반영하였다.

그 결과 SHAP 를 사용하여 속성값을 원본데이터셋의 2/3 로 줄였음에도 불구하고 모델의 F1-score 는 약 0.26 에서 약 0.8 로 오히려 개선되었으며 모델의 학습 시간 또한 감소하였음을 확인하였다. 향후 연구에서는 비정상 거래의 최신 패턴 양상을 반영하여 특이치를 탐지할 수 있도록 하는 연구를 진행할 것이다.

사사문구

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1047113)

참고문헌

[1]김상민, [금융 NCS 공부합시다] 신용카드 부정사용, 한국경제 생글생글, 2018.05.28.

[2]Kaggle, “Credit Card Fraud Detection,” https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[3]손민재, "Conditional GAN 을 활용한 오버샘플링 기법," 한국정보처리학회 2018 년도 추계학술발표대회 2018 Oct. 31, vol.25, no.2, pp.609-612, 2018.

[4]X. Man and Ernest P. Chan, "The best way to select features? comparing mda, lime, and shap," The Journal of Financial Data Science, vol.3 no.1, pp.127-139, 2021.