비지도학습 딥러닝을 활용한 이상거래탐지 시스템 모델

김주현*0, 원정임**
*0SK 주식회사 C&C
**한림학교 스마트컴퓨팅연구소
moonbuggy@sk.com, jiwon@hallym.ac.kr

Fraud Detection System Model Using Unsupervised Deep Learning

JuHyun Kim*0, Junglm Won**
*0C&C, SK Holdings Co.
**Smart Computing Lab., Hallym University

요 약

최근 몇 년간 대량의 데이터를 학습하여 예측·분류의 정확도를 향상시키는 딥러닝 기술이 비약적으로 발전하면서, 전 금융권에서도 이를 실제 업무 프로세스에 활용하는 방안에 대한 관심도가 꾸준히 높아지고 있는 추세이다. 특히, 딥러닝 기술을 활용한 FDS 시스템을 통해 이상거래 탐지율 향상, 신속한 대응, 자동화된 탐지 결과 반영 등을 기대하고 있다. 일반적인 접근방식은 사전에 목표 변수 분류가 이루어진 충분히 많은 훈련 데이터를 이용하는 지도 학습이다. 이런 지도 학습은 사전 데이터 분류에 많은 노력이 요구되고, 보통 현실에서는 비정상 관측 데이터가 거의 없는 경우가 많기 때문에 근본적인 한계를 갖고 있다. 본 논문에서는 현재 운영 중인 일반적인 형태의 FDS 시스템을 분석하고, 지도 학습의 한계를 극복하기 위해 비지도학습인 오토인코더 신경망을 활용한 이상거래 탐지 모델 적용 방안을 제시한다.

1. 서 론

전자금융거래의 거래량 증가와 간편 결제 수단의 다양화, 동시에 나날이 치밀해져가는 사이버 금융사기(Fraud) 방법들은 금융 회사의 이상거래 탐지를 어렵게 만들고 있다. 이에 금융권에서는 전자금융거래를 노린 사기 행위를 사전에 탐지 및 차단하여 거래의 보안성을 향상시키는 이상거래탐지시스템(Fraud Detection System, FDS)을 도입하고 있다 [1]. 한편, 최근 금융권에서도 딥러닝 기술을 활용한 FDS 시스템을 도입하여 이상거래 탐지 및 신속한 대응, 자동화된 탐지 등을 실제 업무 프로세스에 적용하기 위한 방안에 대한 관심이 높아지고 있다. 이를 위한 일반적인 접근방식은 사전에 목표 변수 분류가 이루어진 충분히 많은 훈련 데이터를 이용하는 지도 학습이다. 그러나, 지도 학습은 사전 데이터 분류에 많은 노력이 요구되고, 보통 현실에서는 비정상 관측 데이터가 거의 없는 경우가 많기 때문에 근본적인 한계를 갖고 있다.

본 논문에서는 현재 운영 중인 일반적인 형태의 FDS 시스템을 분석하고, 지도 학습의 한계를 극복하기 위해 오토인코더(Autoencoder) 신경망을 활용한 이상거래 탐지 모델 적용 방안을 제시한다.

2. 현재 구축 상황

2.1. FDS 구성

국내 주요 금융 거래 절차는 그림 1과 같이『이용자 인증 → 거래지시 → 거래 확정』순서로 이루어지며, 이 과정에서 발생하는 정보를 수집하여 이상금융거래 탐지를 수행 한다.

(1) 수집시스템

금융거래 이용자의 PC, 모바일 등에서 수집한 정보 - 단말기, 로그인, 거래정보 등 - 을 축적하고, 분석시스템에서 필요한 형태로 전처리를 수행한다.

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2017R1D1A1B03030969).

(2) 분석시스템

전처리된 정보와 이상금융거래 유형 정보를 분석하여 이상거래 여부를 판정하고, 실제 탐지모델을 구현한다.

(3) 대응시스템

이상거래로 판별된 거래정보에 대해 인증/거부 등을 수행한다.

(4) 모니터링 및 감사

FDS 전 과정에 대한 관리 및 모니터링을 수행한다.

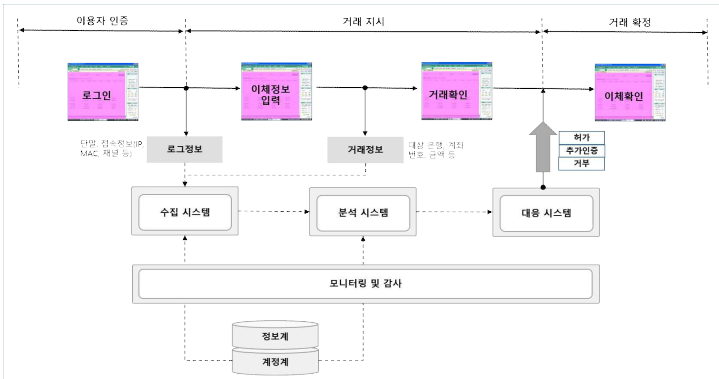

그림 1. 금융거래와 FDS

2.2. FDS 탐지모델

FDS 탐지모델은 크게 오용탐지모델과 이상탐지모델로 구분될 수 있다. 오용탐지모델은 과거의 부정행위패턴을 기반으로 현재 알려진 패턴과 일치하는지 검사하여 부정행위를 탐지하며, 과거정보(사고 정보 등)에 의존하는 방식이다. 이상탐지모델은 정상 금융거래 행위(데이터)를 기준으로 상대적으로 급격한 변화를 일으키거나 확률적으로 낮은 행위가 발생할 경우를 탐지하는 것을 기본 개념으로 한다. 알려지지 않은 부정거래행위에 대한 사전 탐지가 가능하지만, 오탐률이 높은 경향이 있다.

현재 보편적으로 룰 기반의 오용탐지모델이 사용되고 있으나, 알려지지 않은 사고를 예방하는 목적에는 다소

불충분하고, 미탐률이 상대적으로 높다. 또한, 일부 금융회사는 전통적인 머신러닝 기법(예를 들어, 서포트 벡터 머신, 랜덤 포리스트 등)과 딥러닝을 채택해 FDS를 구축 및 운영 중인 것으로 보고되고 있으나, 관련된 제반 사항을 공개하고 있지 않는 실정이다.

3. 오토인코더 신경망 활용 모델 제안

3.1. 모델 설계

본 논문에서 제안하는 오토인코더 신경망 모델에 대한 비교 평가를 위해서 전통적인 지도학습 방법인 심층 신경망(Deep Neural Network, DNN)을 함께 구현하였다.

(1) 심층 신경망(Deep Neural Network)

심층 신경망 모델은 입력 데이터를 비선형적 연산을 통해 재구성하고 네트워크를 학습시키는 딥러닝의 일반적인 방식으로서 다양한 분야에서 적용되고 있다.

본 논문에서 사용한 모델은 그림 2와 같이 191개의 유닛을 포함하는 5개의 숨겨진 층으로 구성되었다. 유닛의 수와 층의 수에 따라 손실 값과 정확도에 차이가 있지만 다양한 형태로 구조를 변경하며 실험한 결과를 토대로, 비교적 높은 정확도를 보이는 구조를 채택하였다.

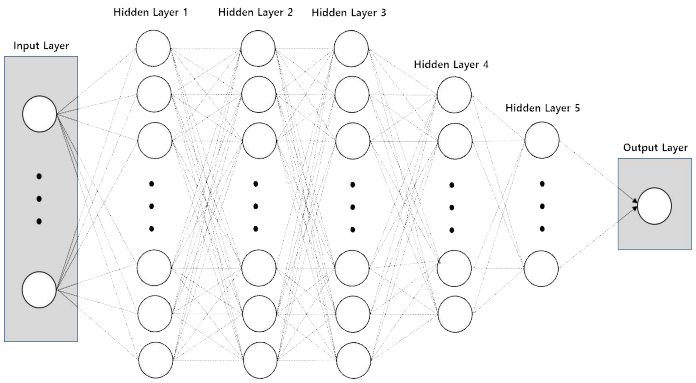

그림 2. 심층 신경망

입력 층과 숨겨진 층간의 활성 함수로는 Relu를, 숨겨진 층과 출력 층간에는 Sigmoid를 적용하였다. 또한 과적합(Overfitting) 문제를 개선하는 드롭아웃(Dropout)을 0.2의 비율로 설정하였다. 학습과정에서는 32의 배치 사이즈를 설정하여 손실 함수의 기울기 하강에 영향을 주어 최저점을 지속적으로 찾을 수 있게 하였으며, 실제 정답과 추정된 정답의 차이로 발생되는 손실 값이 5번의 순환동안 변화가 없다면 최저점에 수렴한 것으로 가정하고 학습을 멈추는 Early Stopping을 적용하였다.

(2) 오토인코더(Autoencoder)

현실 세계에서 비정상 거래 데이터는 매우 적다. 이로 인해 지도학습 방식으로 모델 형성 시 왜곡된 형태가 될 수 있는 가능성이 존재하기 때문에 잘못된 결과가 산출될 수 있다. 오토인코더는 X 클래스(정상 데이터)에 대해 X를 생성하는 모델인데, 정상 데이터만 이용해 모델을 생성한다면 Y 클래스(비정상 데이터)가 들어왔을 때 X를 생성하지 못하게 된다. 본 논문은 이러한 특성을 고려해 오토인코더 신경망을 활용한다. 사용된 모델은 그림 3과 같이 212개 유닛을 포함하는 적층 CAE(Contractive Auto Encoder)로 구성되었다. 심층 신경망과 같이 다양한 형태로 구조를 변경하며 실험한 결과를 토대로, 높은 정확도를 보이는 구조를 채택하였다. 입력 층과 숨겨진 층간의 활성 함수로는 Sigmoid를 적용하였고, 숨겨진 층과 출력 층간에는 Linear를 적용하였다. 학습과정에서는 24의 배치 사이즈를 설정하였고, 심층 신경망과 동일하게 학습을 멈추기 위해서는 Early Stopping을 적용하였다.

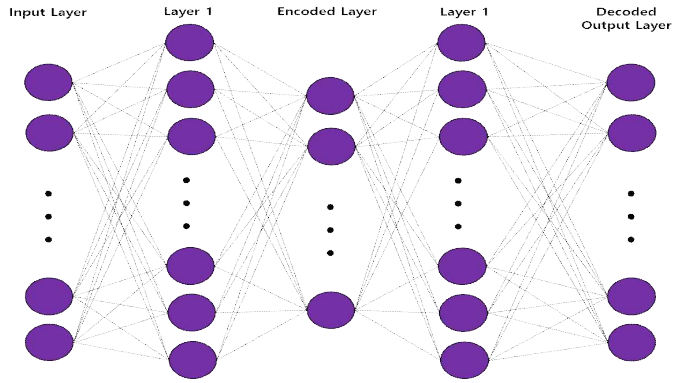

그림 3. 적층 오토인코더

오토인코더는 훈련 집합에 있거나 훈련 집합에 가까운 샘플은 제대로 복원하지만, 훈련 집합에 속하지 않은 데이터, 즉 데이터를 생성한 확률 밀도에 대해 발생 확률이 아주 낮은 샘플은 제대로 복원하지 못한다. 또한, 오토인코더는 매니폴드(manifold) 근방의 데이터에 대해서만 제대로 작동하는 것으로 알려져 있다 [2,4].

본 논문은 정상과 비정상 거래간의 데이터 분포에 차이가 있고, 정상 거래만을 이용해 학습한 모델은 비정상 거래가 입력될 때는 제대로 복원하지 못할 것이라는 가정 하에 모델을 정의 및 훈련시켰다. 한편, 입력 값과 출력 값 간의 차이는 다음과 같이 Mean Square Error(L_i)로 측정했으며, 특정 임계 값을 기준으로 훈련된 데이터가 제대로 복원되었는지 여부를 판정하였다.

$$L_i(x) = (x - G(x))^{\frac{1}{2}}$$

$$Y_i = \begin{cases} 1, L_i(x) \geq 0.01 \\ 0, L_i(x) < 0.01 \end{cases}$$

x는 입력 벡터를 의미하고, G(x)는 입력 벡터에 대응되는 출력 값을 의미한다. Y_i는 이상거래에 대한 판별 값이다. 임계 값은 실제 데이터 분포에 따라 달라지며, 본 논문에서는 몇 차례의 실험 및 결과 산포도를 통해 데이터 복원이 제대로 된 경우와 그렇지 않은 경우를 최대한 구분할 수 있는 경계인 0.01을 임계 값으로 설정하였다. (참고: 그림 4. 이상 거래 데이터의 입출력 차이)

3.2. 데이터 준비

(1) 데이터 정의

실제 금융기관에 적용된 이상거래탐지시스템에서 수집하는 50여개 항목 중 이상탐지모델에 활용하기에 적합한 14개 항목을 표 1과 같이 정의한다. 표 1에 포함되지 않은 항목, 예를 들어 Black List 포함 대상, 단말고유번호 등은 오용탐지모델에서 사용되는 것들로 정형화된 룰에 적합한 정보이므로, 고려 대상에서 제외한다.

(2) 기초 데이터 작업

개인정보보호법 등과 같은 법적 규제로 인하여 실제 데이터를 사용할 수 없는 관계로, 다음과 같은 방식으로 가공의 데이터를 생성하였다. 사기 거래에 취약한 계층인 노년층을 기준 집단으로 하고, 일반적인 금융 거래 내역과 유사하게 임의적으로 정상 거래 데이터를 생성하였고, 금융감독원 소비자 경보에 소개된 사기 거래 사례를 토대로 이상 거래 데이터를 생성하였다. 예를 들어, 금융사기가 증가하는 추석명절 이전, 성별 및 나이 등에 따른 금융사기 유형, 금융사기 발생 시간대, 접속 채널 유형, 접속 국가 등을 고려하였다 [3].

표 1. FDS 수집 정보

수집대상	수집정보	비고
단말정보	매체유형	인터넷, 모바일, ATM 등
접속정보	접속국가	국가별로 위험 구분
고객정보	성별	남, 여 구분
	생년월일	연령계산에 활용
계좌정보	개설일	기간 단위로 범주화
	최종거래일	기간 단위로 범주화
	상품유형	보통예금 등
금융거래정보	수취계좌의 해외계좌 여부	Y, N 구분
	업무구분	타행이체, 대출 등
	거래금액	일정 범위 단위로 범주화
	평균금액	평균거래금액
	거래시간	시간대 별로 위험 구분
	거래주기	최종거래 이후 경과 일

4. 실험 및 성능 평가

4.1. 실험환경

본 연구에서는 케라스 프레임워크를 이용하여 딥러닝 모델을 구현하였다 [5]. 윈도우즈 10 환경에 케라스, 텐서플로우 1.6 버전을 설치하였고, Jupyther notebook과 Python을 이용해 개발하였다.

4.2. 학습 및 테스트

표 2와 같이 5만 건의 정상 거래, 5천 건의 이상 거래 데이터를 이용해 각각 심층 및 오토인코더 신경망을 훈련시켰다. 훈련 시 심층 신경망은 정상 및 비정상 거래 데이터가 혼합되었으나, 오토인코더는 정상 거래 데이터만 사용하였다. 훈련이 완료된 이후 테스트 데이터를 이용해 정상/비정상 거래 판별 여부를 측정하였다. 두 신경망 간의 정확도 차이는 표 3과 같다.

표 2. 데이터 분류

분류	심층신경망(건)	오토인코더(건)
학습	정상 4만, 비정상 4천	정상 4만
테스트	정상 1만, 비정상 1천	정상 1만, 비정상 5천

표 3. 판별 정확도

구분	정확도(%)	비고
심층신경망	99.92	표2의 테스트 데이터 대상으로 판별 정확도 측정
오토인코더	95.11	

그림 4는 오토인코더 신경망 대상으로 5천 건의 비정상 거래 데이터를 입력했을 때 각 데이터별 입출력 값의 차이를 측정한 결과를 보여준다. 대부분의 경우에서 출력 값이 제대로 복원되지 못함을 확인할 수 있다. 한편, 그림에서 점선은 제대로 복원된 경우와 그렇지 않은 경우를 구분 짓는 명확한 경계가 된다.

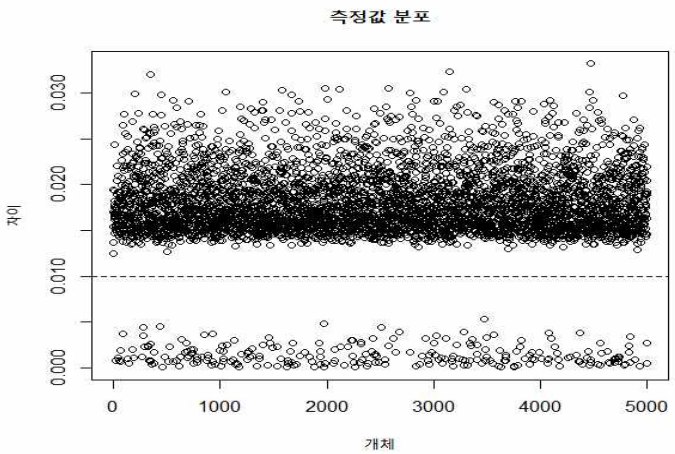

그림 4. 이상 거래 데이터의 입출력 차이

4.3. 평가

지도학습으로 구현된 일반적인 심층신경망 모델이 정확도가 다소 높았으나, 비지도학습인 오토인코더 모델을 이용해도 높은 정확도를 보임을 확인할 수 있다. 이상 거래 데이터를 획득하는 것이 매우 힘든 상황을 감안할 때, 실제 비즈니스 현장에서는 오토인코더 모델을 적용하는 것을 적극적으로 검토해 볼 수 있다고 판단된다.

5. 결론 및 향후 연구과제

본 연구에서는 현재 대부분의 금융 회사에서 구축 운영 중인 이상거래시스템의 한계로 인식되고 있는 이상탐지모델을 개선하기 위해 비지도학습의 오토인코더 신경망을 활용한 방안을 제시하였다. 현실적인 제약으로 실제 거래 데이터를 활용하지 못한 한계가 존재하나, 금융거래 특성을 반영하여 생성된 거래 데이터를 사용하여 충분히 제안한 모델의 유용성을 보였다. 향후, 기존 룰 기반의 오용탐지 모델과 제안 모델의 결합 및 금융 거래 순서 반영을 통한 전체 이상거래 탐지 시스템의 고도화뿐만 아니라 상시감사, 부도예측, 금융시장 분석, 재무제표 분석 등과 같은 금융의 다양한 영역으로 딥러닝 기술을 확대 적용하고자 한다.

참고문헌

[1] 금융보안원 “머신러닝 기반의 이상거래 탐지 시스템 동향”, 정보보호동향, 2017.

[2] Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning MIT Press, 2016.

[3] 금융감독원 소비자경보, [Online]. Availalbe: http://consumer.fss.or.kr

[4] 오일석, 기계학습, 한빛아카데미, 2017.

[5] The Keras Blog [Online]. Availalbe: https://blog.keras.io/index.html