Information Systems Review
제22권 제1호

ISSN : 1229-5078(Print) 2713-8143(Online)

생성적 적대 신경망과 딥러닝을 활용한 이상거래탐지 시스템 모형

김예원, 유예림, 최홍용

To cite this article : 김예원, 유예림, 최홍용 (2020) 생성적 적대 신경망과 딥러닝을 활용한 이상거래탐지 시스템 모형 , Information Systems Review, 22:1, 59-72

Information Systems Review
Vol. 22, No. 1 February 2020 http://dx.doi.org/10.14329/isr.2020.22.1.059

생성적 적대 신경망과 딥러닝을 활용한 이상거래탐지 시스템 모형

Fraud Detection System Model Using Generative Adversarial Networks and Deep Learning

김 예 원 (Ye Won Kim) LG CNS 금융/공공빅데이터팀 사원
유 예 림 (Ye Lim Yu) LG CNS 금융/공공빅데이터팀 사원
최 홍 용 (Hong Yong Choi) LG CNS 금융/공공빅데이터팀 팀장, 교신저자

요 약

인공지능이 다루기 어려운 개념에서 아주 익숙한 도구로 자리매김 하고 있다. 이와 더불어 금융권에서도 인공지능 기술을 도입하여 기존 시스템의 문제점을 개선하고자 하는 추세이며, 그 대표적인 예가 이상거래탐지 시스템(Fraud Detection System, FDS)이다. 결제 수단의 다양화 및 전자금융거래의 증가에 따라 치밀해져 가는 사이버 금융사기(Fraud)를 기존의 규칙기반 FDS로는 탐지하기 어려워지고 있다. 이를 극복하기 위해 딥러닝 기술을 적용하여 이상거래 탐지율을 향상시키고, 이상행위에 즉각 대응하며, 탐지 결과의 반영을 자동화하고자 하는 시도가 이루어지고 있다. 딥러닝 FDS 구축에서 핵심 문제는 데이터 불균형과 이상거래 패턴의 변동이다. 본 논문에서는 생성적 적대 신경망(Generative Adversarial Network, GAN)을 활용한 오버샘플링 기법을 통해 데이터 불균형 문제를 개선하고, 이상거래 분류기로써 심층 신경망(Deep Neural Network, DNN)과 합성곱 신경망(Convolutional Neural Network, CNN)을 적용하여 이러한 문제를 개선하고자 하였다. 실험 결과, GAN 오버샘플링이 이상거래 데이터의 불균형 문제를 개선하는데 효과를 보였으며, WGAN이 가장 높은 개선 효과가 있음을 확인하였다. 또한 제안 FDS 모형의 AUC가 0.9857로 랜덤포레스트 FDS 모형에 비해 약 6.5% 향상되어, 딥러닝이 이상거래 탐지에 뛰어난 성능을 가짐을 입증하였다. 더불어 딥러닝 모형 중 DNN은 CNN에 비해 오버샘플링의 효과를 더 잘 반영함을 확인하였다.

키워드 : 이상거래탐지, 딥러닝, 생성적 적대 신경망

Ⅰ. 서 론

금융거래 기술의 발전과 함께 금융거래 사기 수법 또한 더욱 치밀하고 다양해지며 이상거래 탐지시스템(Fraud Detection Model, FDS)은 금융기업의 필수 시스템으로 자리매김하였다. FDS는 전자 금융거래 발생 시, 접속 정보와 거래 정보 등을 수집하고 분석하여 이상금융거래 유무를 판별하고 그에 대응하는 복합적인 시스템이다.

초기의 FDS는 이상거래 분석/탐지 과정에서 규칙기반의 탐지시스템을 적용한 것으로, 이상거래에 해당하는 규칙을 정의하고 이를 기반으로 이상거래와 정상거래를 판별한다. 규칙기반 FDS는 속도가 빠르며 개발이 쉽다는 장점이 있어 많은 금융 기업에서 사용해왔다. 그러나 규칙에 포함되지 않은 거래 패턴에 대해서는 이상거래 여부를 판별할 수 없다는 한계점을 가지고 있다. 최근 거래량 및 거래과정의 복잡도 증가와 더불어 이상거래의 패턴이 다양해지면서, 이러한 한계점을 보완할 수 있는 효율적이고 고도화된 FDS에 대한 요구가 높아지고 있다. 이와 같은 요구는 머신러닝과 딥러닝 기법을 적용한 FDS 연구로 이어졌다. 본 연구에서는 FDS에 GAN과 딥러닝 기법을 적용하여, 규칙기반 FDS의 한계를 극복함과 동시에 탐지 성능을 향상시키고자 한다.

즉, 본 연구에서는 다양한 GAN 기법과 딥러닝 분류 기법의 조합을 통해 보다 향상된 성능의 이상거래탐지 모형을 제안하고자 한다.

Ⅱ. 연구 배경 및 기존 연구

현재까지 FDS 개발에서 기계학습 기법을 통해 개선하고자 하는 중요 문제로 “데이터 처리 속도”, “데이터 불균형”, “이상거래 패턴의 불규칙한 변동” 등이 있다. 데이터 처리 속도는 실시간 대응을 필요로 하는 FDS의 특성으로 인해 제기되는 문제며, 데이터 불균형과 이상거래 패턴 변동의 경우 분석/탐지 모델링 영역에서 해결해야 하는 주요 문제라고 할 수 있다.

2.1 데이터 처리 속도

금융 거래, 특히 카드 거래의 경우 거래의 발생과 이로 인한 결과가 실시간으로 반영되기 때문에 이상 거래에 대한 탐지가 즉각적으로 이루어져야 한다. 따라서 데이터 처리 속도는 FDS의 핵심 요건 중 하나이다. 기존의 전통적인 통계기법 및 데이터 수집 기법은 이러한 요건을 만족시키기 쉽지 않아 FDS의 데이터 처리 속도 문제를 머신러닝과 딥러닝 기법을 통해 개선하고자 하는 시도가 다수 이루어지고 있다(박재훈 등, 2015; 이용현 등, 2019).

데이터 처리 속도의 향상을 위한 가장 직관적인 방법은 일반적으로 데이터의 차원을 축소시키는 것이다. 그러나 차원 축소는 중요한 정보의 손실을 야기할 수 있기 때문에, 이를 해결하기 위한 효율적인 차원 축소 기법이 연구되어 왔다. 차원 축소를 위한 가장 전통적인 통계 기법은 주성분 분석(Principal Component Analysis, PCA)이다. 주성분 분석은 직교 변환을 통해 고차원의 데이터를 저차원의 데이터로 환원시키는 차원 축소 기법으로, 가장 널리 사용되지만 선형 관계만 모델링 가능하다는 한계가 존재한다.

이용현 등(2019)은 이러한 주성분 분석의 한계를 보완하는 차원축소 기법으로 오토인코더(AutoEncoder)를 제안하였다. 오토인코더는 출력이 입력과 동일해지도록 하는 것을 목표로 하는 인공신경망 기법이다. 즉, 입력 데이터가 은닉층을 거쳐 부호화(encode)된 후 다시 복호화(decode)하였을 때 입력을 충실히 재현할 수 있는 데이터의 특성을 찾아내는 것이다. 오토인코더는 규제 기법의 적용과 은닉층의 변형을 통해 필요에 따라 다양한 구성이 가능하여 변수들 간의 비선형 관계 파악에 용이하다. 이용현 등(2019)에 따르면 오토인코더의 차원 축소 기능을 FDS에서 전처리 기로 사용할 경우 정밀도와 재현율의 큰 손실 없이 새로운 인스턴스의 분류 속도를 약 10% 정도 빠르게 할 수 있음을 보였다.

차원 축소 외에도 데이터 처리 과정에서의 효율성을 높이고자 다양한 연구가 시도되었다. 박재훈 등(2015)은 의사결정나무를 통해 이상거래탐지 규칙을 정규화 하여 효율성을 높이고자 하였고, 이러한 방식이 데이터 처리 속도를 높이고 처리비용을 절감하는 효과가 있음을 보여주었다.

2.2 데이터 클래스 불균형

FDS와 같은 이상징후 탐지에 사용되는 데이터의 경우, 대부분 이상 데이터 건수가 정상 건수에 비해 현저히 적은 불균형 데이터이다. 이와 같은 클래스 불균형은 모형 학습에서의 편향, 성능 평가 등 다양한 문제를 발생시킨다(Burez and Van Den Poel, 2009).

소수 클래스가 절대적으로 적은 수를 가지거나, 다수 클래스에 비해 상대적으로 적은 비율일 때 이를 해결하는 보편적인 방법은 샘플링(Sampling) 기법을 통해 데이터 클래스의 비율을 맞추는 것이다(Burez and Van Den Poel, 2009). 그 예로, 소수 클래스를 늘려 균형을 맞추는 오버 샘플링, 다수 클래스를 제거하는 언더 샘플링, 오버 샘플링과 언더 샘플링을 합성하여 만든 Synthetic Minority Over-sampling Technique(SMOTE)가 있다(Chawla *et al.*, 2002; 손민재 등, 2018). 김량형 등(2016)은 기업부실화 예측 모형 개발 시 발생되는 데이터 불균형 문제를 해결하기 위해 SMOTE를 적용하였으며 예측성과가 향상됨을 확인하였다. 언더 샘플링은 결과적으로 샘플 수가 줄어들기 때문에 정보의 손실을 발생시킬 수 있다는 단점이 있다(김한용, 이우주, 2017). 오버 샘플링은 정보의 손실은 발생하지 않으나 데이터의 수가 늘어나 모형 구축에 시간이 오래 걸리며 과적합 가능성이 있다(He and Garcia, 2009). 최근에는 GAN을 활용한 오버샘플링 기법이 다수 제안되었는데(Douzas and Bacao, 2018; 서상현 등, 2017; 손민재 등, 2019), 데이터의 특성 및 분석 목적에 따라 다양한 변형 모형을 적용하여 위조 데이터를 생성할 수 있다는 장점이 있다.

샘플링 외에 클래스 불균형을 해결하는 방법에는 하나의 클래스만을 이용하여 분류 모형을 학습시키는 OCC 기법(One-Class Classification)이 있다. OCC 기법은 모형에 하나의 클래스만을 학습시킨 후 해당 클래스인지 또는 그 외 클래스인지 예측하는 기법이다. Zheng *et al.*(2018)은 LSTM-오토인코더를 통해 정상 거래 패턴을 학습하고 Complementary GAN을 통해 그 패턴에 반대되는 위조 데이터를 생성하여 "정상" 패턴과 그 외의 패턴을 구분하는 2단계로 구성된 OCC 기반의 FDS 모형을 제안하였다.

2.3 이상거래 패턴의 변동

앞서 언급하였듯이, 규칙기반 FDS의 가장 큰 한계점은 새롭게 변동하는 이상거래 패턴에 대한 탐지가 불가능하다는 것이다. 이를 극복하고자 머신러닝과 딥러닝 기법을 적용한 FDS 연구가 진행되고 있다.

대부분의 연구에서 FDS 분류기 구축 시 지도학습 기법을 사용하였다. 손민재 등(2018)은 conditional GAN(cGAN)을 통해 오버 샘플링된 데이터에 서포트벡터머신(Support Vector Machine)(Hearst *et al.*, 1998), 랜덤포레스트(Random Forest)(Ham *et al.*, 2005; Liaw and Wiener, 2002), 다중 퍼셉트론(Multilayer Perceptron) (Haykin, 2009) 세 종류의 지도학습 기법을 적용한 분류기를 구축한 후 각 분류기별 성능을 비교하였다. Purushu *et al.*(2018)은 Spark ML과 Azure ML을 통해 랜덤포레스트, 서포트벡터머신, 로지스틱회귀모형 등의 성능을 비교한 결과 랜덤포레스트의 성능이 가장 뛰어남을 확인하였다. 김주현, 원정임(2018)은 오토인코더를 활용한 비지도학습기반 분류기를 구축한 후, 이를 심층신경망(Deep Neural Network, DNN) 분류기와 비교 평가하였다. 비지도학습 기반의 분류기는 오토인코더에 정상 거래 데이터만을 학습시키면 이상 거래가 입력될 때 제대로 복원할 수 없을 것이라는 가정 하에 수립되었다. 연구 결과, 지도학습 분류기의 성능(99.92%)이 다소 높으나 비지도학습분류기 또한 높은 성능(95.11%)을 가졌음을 보여주었다.

Ⅲ. 이론적 배경

제Ⅲ장에서는 제Ⅱ장에서 소개한 문제 중 데

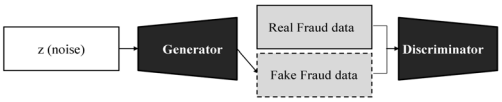
이터 클래스 불균형과 이상거래 패턴 변동에 집중하여 논의를 진행하여, 이러한 문제를 해결하는데 적용할 딥러닝 기법과 본 연구가 제안하는 FDS 모형의 구성을 소개한다.

3.1 GAN 기법을 이용한 오버샘플링

3.1.1 (Vanilla) GAN

GAN은 2014년 Ian Goodfellow가 처음 발표한 개념으로, 실제 데이터의 분포를 따라 위조 데이터를 만드는 생성모형(Generator)과 실제 데이터와 위조 데이터를 구분하는 판별모형(Discriminator)이 서로 대립하여 학습하는 구조이며 각각의 모형은 신경망으로 이루어져 있다(Goodfellow *et al.*, 2014). G를 생성모형, D를 판별모형이라 할 때, P_data는 실제 데이터의 분포를 의미하고, P_z는 잡음(noise) 혹은 잠재 변수라고 불리는 z의 분포이며, P_g는 생성모형에서 만들어진 데이터의 분포이다. 이때 판별모형 $D(x)$는 x가 생성모형의 분포 P_g가 아닌, 실제 데이터의 분포 P_{data}를 따를 확률을 의미한다. 최소최대 의사결정 이론(Minimax Theory)으로 G와 D의 관계를 손실함수로 표현하면 아래 식 (1)과 같으며 G는 이 값을 최소화하는 방향으로, D는 최대화하는 방향으로 학습하게 된다.

$$min_g max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_{Z(z)}}[\log(1-D(G(z)))] \quad (1)$$

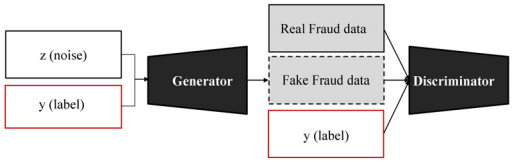


〈그림 1〉 GAN을 활용한 오버샘플링 흐름도

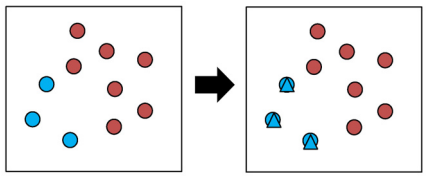
3.1.2 Conditional GAN(cGAN)

기본GAN(vanilla GAN)을 사용하면 위조하는 데이터에 대한 제약(또는 조건부여)이 불가능하다. cGAN은 이 점을 보완하기 위해 제안된 기법으로 Vanilla GAN과 유사한 학습 방식을 가지지만 생성모형의 입력 변수에 차이가 있다(Mirza and Osindero, 2014). cGAN은 입력 변수에 잡음 z값뿐만 아니라, 생성모형과 판별모형에 특정 조건을 나타내는 y값이 추가된 구조가 특징이다(식 (2), <그림 2>). 여기서 y값은 라벨(label)로 cGAN이 위조 데이터를 생성-판별하는 과정에서 참조하는 추가 정보이다. 예를 들어, K-평균 군집(K-means Clustering)을 통해 군집 라벨을 부여하면 위조 데이터 생성-판별 시 입력데이터의 군집을 고려하게 되는 것이다. 라벨 y는 분석가의 재량에 따라 다양한 형태로 구성될 수 있다. <그림 2>에서는 위조하고자 하는 데이터 이자소수 클래스인 이상거래(Fraud)에 대해 라벨 y를 부여한 cGAN 구조를 나타내었다.

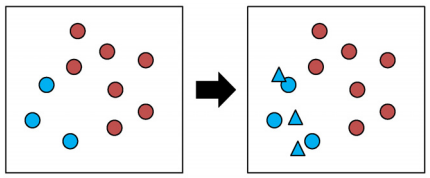
$$min_g max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x|y)] + E_{z\sim P_{Z(z)}}[\log(1-D(G(z|y)))] \quad (2)$$



〈그림 2〉 cGAN을 활용한 오버샘플링 흐름도



〈그림 3〉 단순 오버샘플링 기법



〈그림 4〉 GAN 오버샘플링 기법

3.1.3 Wasserstein GAN

기계학습 분야에서 가장 중요한 과정 중 하나는

확률분포를 학습하는 과정이다. 데이터의 결합 확률 분포를 학습하는 것은 생성모형을 만드는 과정에 있어 필수적 과정이다. 이는 최대 우도 함수 추정(Maximum Likelihood Estimation)은 Kullback-Leibler Divergence(KLD)를 최소화하는 과정과 일치한다(식 (3)). KLD는 이상적인 분포를 기준으로 근사 분포를 생성할 때 발생할 수 있는 정보 엔트로피 차이를 계산하는 과정이다. 또한 KLD는 Jensen Shannon Divergence(JSD)로 나타낼 수 있음이 증명되면서(Murphy, 2012), 이상적인 분포와 근사 분포간의 차이를 줄이는 수학적인 접근으로 사용되고 있다.

KLD와 JSD의 정의에 따르면, 이상적인 분포와 근사 분포는 같은 서포트(Support)를 가져야 한다. 여기서 서포트란 확률 변수가 취할 수 있는 모든 수들의 집합을 의미한다. 그러나 실제 데이터에서 유의미한 데이터를 가진 매니폴드(Manifold)는 데이터 전체 공간에 비해 상대적으로 작은 공간에 밀집되어 있기 때문에 비교하는 두 분포의 서포트는 같지 않을 확률이 높다. 따라서 손실함수가 생성모형을 통해 만들어진 위조 데이터의 분포와 실제 분포간의 거리를 제대로 계산하지 못하게 되어 GAN의 학습이 어렵게 되는 것이다. 이 점을 보완하고자 두 확률 분포간의 거리를 계산하는 측도로 Wasserstein Distance가 제안되어 기존의 GAN이나 cGAN의 손실함수로 적용하게 되었고 이를 Wasserstein GAN(WGAN), Wasserstein Conditional GAN(WCGAN)이라 칭한다(Arjovsky *et al.*, 2017).

Wasserstein Distance는 이상적인 분포의 모양으로 근사 분포를 일치시키기 위해 확률 질량 함수(Probability Mass Function)에서의 질량(mass)을 옮기는 과정에서, 손실이 최소화되는 방법을 채택하였을 때 두 확률 분포 간의 거리를 계산한다.

$$\begin{aligned}\mathrm{agmax}\sum_{i=1}^{n}\log P_\theta(x_i) &= \int_x P_r(x)\log P_\theta(x)dx \\ &= \mathrm{argmin}_\theta KL(P_r\|P_\theta)\end{aligned} \quad (3)$$

x를 P_r의 서포트에 있는 막대(bin) 중 한 개, y를 P_θ의 서포트에 있는 막대 중 한 개라고 할 때, $\gamma(x,y)$ 는 x와 y의 거리, 즉 질량을 의미한다. Wasserstein Distance를 도입한 GAN 손실함수는 식 (4)와 같이 정의되며, 이를 최소화하는 방향으로 GAN 학습이 진행된다.

$$\begin{aligned}Cost &= \mathrm{mass}\times\mathrm{distance} \\ &= \sum_{x\in X}\sum_{y\in Y}\gamma(x,y)\cdot\|x-y\|^p \\ &= E_{\gamma(x,y)}(\|x-y\|^p)\end{aligned} \quad (4)$$

3.1.4 GAN 오버샘플링

기존의 오버샘플링 기법은 존재하는 데이터를 그대로 복사하여 해당 데이터에 대한 학습률을 높이게 된다(<그림 3>). 이러한 방법은 분류기로 하여금 동일한 데이터를 여러 번 학습시켜 새로운 정보를 학습시키기 어렵고 과적합 가능성이 높다는 단점이 있다. 반면, GAN 기반의 오버샘플링 기법은 기존 데이터의 분포를 학습하고 해당 분포를 따르는 새로운 데이터를 생성하여 유사하지만 다른 정보를 분류기가 학습할 수 있도록 한다(<그림 4>). 따라서 GAN을 통해 이상거래 데이터를 오버샘플링하면, 기존의 이상거래 건 뿐만 아니라 해당 이상거래 건과 유사한 패턴을 가지는 향후 발생 가능성이 있는 새로운 이상거래 건 또한 학습할 수 있다.

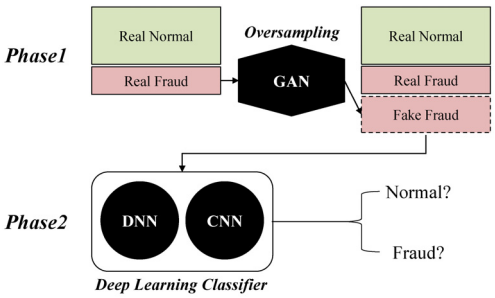
3.2 FDS 모형 구성

본 연구에서 제안하고자 하는 FDS 모형은 데이터 불균형과 이상거래 패턴의 변동 문제에 초점을 맞추어, GAN 오버샘플링-딥러닝 분류기의 두 단계로 구성하였다.

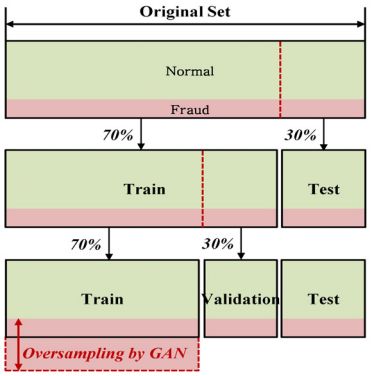
첫 번째 단계에서는 제3.1절에서 소개한 GAN의 생성모형을 통해 이상거래 클래스를 오버샘플링하여 불균형 문제를 처리할 것이다. 이때, GAN 기법은 가장 일반적인 Vanilla GAN 뿐만 아니라

cGAN, WGAN, WCGAN의 세 가지 변형 GAN을 추가 적용하여 각 기법의 오버 샘플링 결과를 확인하고 GAN 오버 샘플링의 정합성을 확인하고자 한다. 사용하여 이상거래 패턴을 군집화한 정보를 추가한다면, 위조 데이터 또한 실제 이상거래의 패턴을 반영하여 생성되어 분류기의 성능을 높일 수 있을 것이라 판단하였다. 또한, WGAN의 경우 Wasserstein Distance를 손실함수에 적용하여 학습력을 높인 알고리즘이기 때문에 GAN에 비해 실제와 유사한 위조 데이터를 생성하여 분류기의 학습 성능이 보다 뛰어날 것으로 기대하여 추가 적용하였다.

불균형 문제가 처리된 데이터에 대해서는 지도학습 딥러닝 기법인 DNN(Hinton *et al.*, 2006)과 CNN(Wang *et al.*, 2014)을 적용하여 전통적인 머신러닝 기법에 비해 높은 분류 성능을 가진 FDS 모형을 개발하고자 한다.



〈그림 5〉 제안 FDS 모형 구조



〈그림 6〉 데이터 세트 분리

Ⅳ. 실험

4.1 데이터 세트 및 실험 환경

4.1.1 데이터 세트 소개

본 실험에서는 세계적으로 유명한 데이터 사이언스 경진 대회인 Kaggle에서 제공하는 신용카드 사용기록(Dal Pozzolo *et al.*, 2015) 데이터 세트를 이용하여 실험을 진행하였다. 해당 데이터는 2013년 9월에 이틀간 기록된 신용카드 사용 내역으로, 보안을 위해 주성분 분석 과정을 거친28개의 주성분 변수와 거래 시간, 금액 변수로 구성되어있다. 또한 총 284,807건의 전체 데이터 중 0.17%인 492건만이 사기 거래로 분류 되어있는 불균형 데이터 세트이다. 본 연구에서는 각 변수의 세부적 의미에 대한 탐구가 아닌 FDS 모형 자체를 구축하는데 초점을 맞추고자, 금액과 거래 시간 변수를 제외하고 주성분 변수28개(V1~V28)와 클래스(이상거래 여부) 변수를 이용하여 실험을 진행하였다.

4.1.2 실험환경

본 연구에서의 실험은 LG CNS의 머신러닝/딥러닝(MLDL) 플랫폼인 Data Analytics & AI Platform (이하 DAP MLDL)에서 진행하였다. DAP MLDL은 Kubernetes 및 Docker Container 기술을 이용하여 포털에서 요청하는 리소스(CPU/GPU/Memory)를 기준으로 서버 장비의 리소스를 일부 할당하여 분석환경을 제공하는 데이터 분석 플랫폼이다. CPU는 E5-2620 v4(2.10GHz 8Core)×2P, 메모리는 32GB DDR4 RAM, 디스크는 1.2TB SSB×2ea (RAID1), OS는 Ubuntu 16.04.5 LTS로 지정하여 사용하였고, DAP MLDL내에서 Python 2.7 Jupyter notebook을 활용하여 작업하였다. 주요 머신러닝/딥러닝 라이브러리인 TensorFlow와 Keras의 버전은 각각 1.11.0, 2.2.4를 사용하였고, Scikit-Learn, Numpy, Pandas와 같은 분석용 라이브러리 등을 활용하였다.

4.2 GAN 오버샘플링

4.2.1 데이터 세트 분리 및 전처리

GAN 오버샘플링 기법을 적용하기에 앞서 28만 건의 전체 데이터를 학습(Train), 검증(Validation), 평가(Test) 데이터로 분할하는 과정을 거쳤다. 학습 데이터는 전체 데이터의 49%, 검증 데이터는 21%, 평가 데이터는 30%의 비율로 나누었다. 학습 데이터와 검증 데이터는 전체 데이터 중 평가 데이터를 제외한 나머지를 7:3의 비율로 나눈 것이다. 그 결과, 학습 데이터 약 14만 건, 검증 데이터 약 6만 건, 평가 데이터 약 8만 건으로 실험을 진행하였다. 오버 샘플링 기법은 검증, 평가 데이터엔 적용하지 않고, 학습 데이터의 이상거래 클래스에 대해서만 진행하였다.

cGAN과 WCGAN의 경우, 소수 클래스에 대한 라벨을 입력 데이터에 추가하는 과정을 거쳤다. 라벨 부여 방식은 분석가의 재량에 따라 다양하나, 본 연구에서는 직관적인 비지도 분류가 가능한 K-평균 알고리즘을 사용하였다.

〈표 1〉 데이터 세트 별 건수

(단위: 건)

		학습 (Train)	검증 (Validation)	평가 (Test)
정상거래		139,319	59,701	85,291
이상 거래	실제	235	109	152
	위조	136,535	-	-
합계		276,089	59,810	85,443

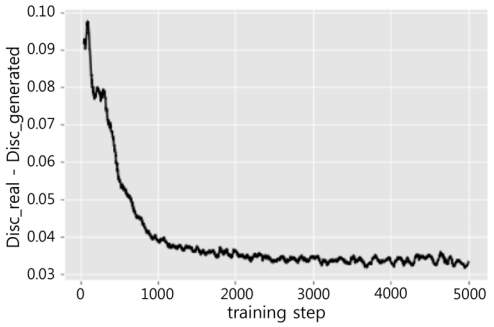
4.2.2 위조 데이터 생성

약 14만 건의 훈련 데이터 중 이상거래로 분류된 240건의 데이터를 이용해 GAN 학습을 진행한 후, 위조데이터를 생성 과정을 진행하였다.

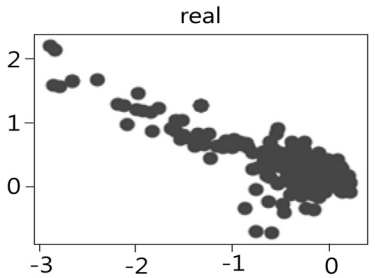
이때 학습 과정에 따른 위조데이터의 형태를 시각적으로 확인하고자 주성분 변수 중 V1을 x축, V2를 y축으로 하는 산점도를 확인하였다(<그림 9>). 산점도를 통해 랜덤한 데이터 뭉치로부터 시작한 위조데이터가 학습을 진행할수록 실제 데이터의 분포 양상(<그림 8>)과 유사해짐을 확인할 수 있었다. 또한 판별모형이 실제 데이터를 실제라고 판별하는 학습 과정 중 생긴 손실과 위조데이터를 위조라고 판별하는 학습과정 중 생긴 손실의 차를 이용하여, 위조 데이터 생성을 위한 최적의 Epoch값과 해당 Epoch의 가중치 값을 구하였다(<표 2>). 이러한 최적 값을 이용하여 훈련 데이터의 정상/이상거래 클래스의 비율이 1:1이 되도록 위조 이상거래 데이터를 14만 건 가량 생성하였다.

〈표 2〉 위조 데이터 생성을 위한 최적 Epoch

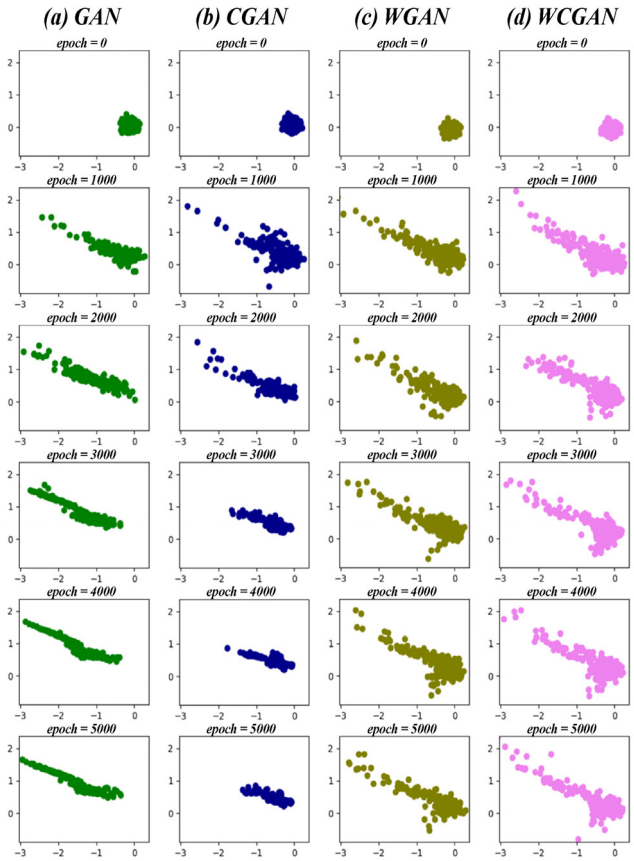
GAN 모형	최적 Epoch
GAN	2,400
cGAN	700
WGAN	4,000
WCGAN	3,100



〈그림 7〉 Epoch 별 WCGAN 판별모델 손실추이



〈그림 8〉 실제 데이터 V1-V2 산점도



〈그림 9〉 Epoch에 따른 위조 데이터 V1-V2 산점도

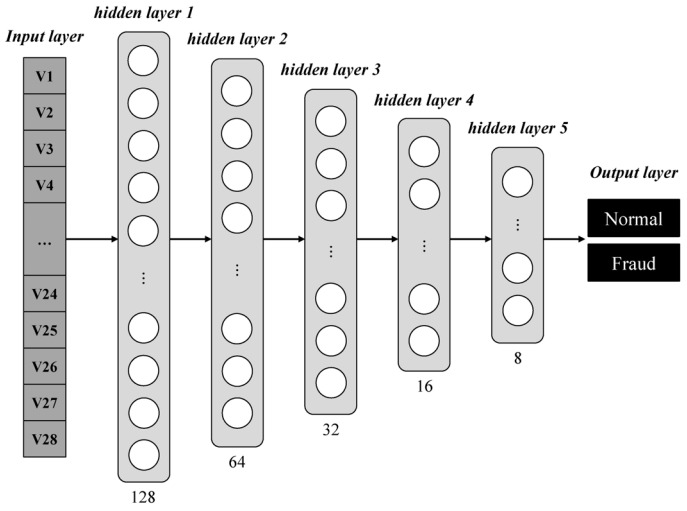
4.3 딥러닝 분류기

4.3.1 DNN 모형

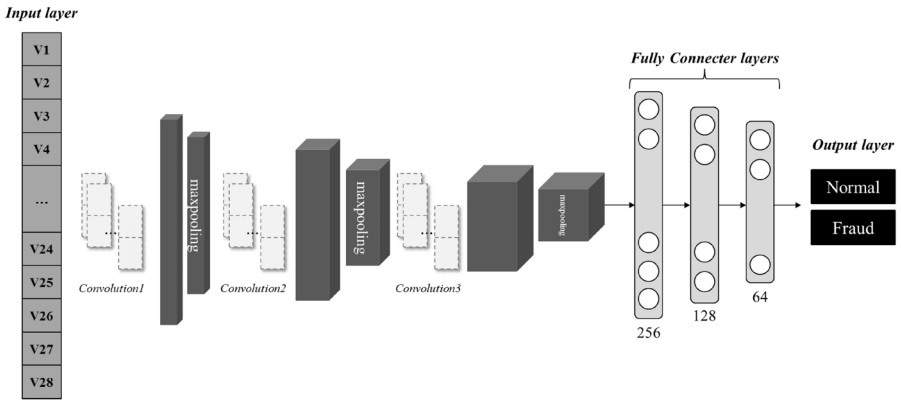
기본적인 딥러닝 모형인 DNN 모형을 활용한 FDS분류기는 총 5개의 은닉층으로 구성하였다. 최종 출력층에는 Softmax 활성화 함수를 이용하여 이진 분류기의 형태로 설정하였고, 그 외의 은닉층은 ReLU 함수를 사용하였다. 또한 과대적합을 방지하고자 각 은닉층에서의 Dropout 비율을 0.3으로 하고, 배치 정규화(Batch Normalization)를 수행하도록 하였다. 최적화 함수로는 Adam을 사용하였고, 학습율은 0.001, 손실함수는 Binary Cross Entropy를 활용하였다. 총 20번의 Epoch을 500 Batch Size로 돌려 학습을 진행하였다(<그림 10>).

4.3.2 CNN 모형

CNN의 경우 입력 데이터가 Convolution-Pooling 과정을 3번 통과한 후, Fully Connected Layer를 3번 통과하여 최종적으로 분류를 진행할 수 있는 구조로 설정하였다. Convolution-Pooling 과정에서 커널의 크기는 2로 통일하였고, 커널 개수는 128 → 64 → 32로 감소하도록 하였다. 또한 DNN과 마찬가지로 과적합 방지를 위해 Dropout(비율 0.2)과 배치 정규화를 수행하였다. Fully Connected Layer는 256 → 128 → 64 순으로 은닉층의 유닛 개수를 설정하였고, Convolution-Pooling 과정과 같이 각 은닉층의 Dropout 비율을 0.2로 하였다. 또한 출력층의 활성화 함수는 Softmax 함수를 사용하였고, 그 외의 층은 모두 ReLU 함수를 사용하였다.



〈그림 10〉 DNN 구조



〈그림 11〉 CNN 구조

최적화 함수에는 학습률을 0.001로 설정하고 Binary Cross Entropy 손실함수를 적용한 Adam을 이용하였다(<그림 11>).

V. 결 과

5.1 실험 결과

본 논문에서 제안한 GAN 오버샘플링 - 딥러닝 분류기 모형과의 비교를 위해 DNN, CNN과 더불어 전통적 머신러닝 기법인 랜덤포레스트(Random Forest, RF)의 성능을 함께 확인하였다. 랜덤포레스트는 121개의 데이터세트에 대한 17가지의 분류 문제에서 179개의 분류 알고리즘을 비교하였을 때, 가장 높은 성능을 보였기 때문에(Fernández-Delgado *et al.*, 2014), 본 실험에서의 베이스 모형으로 지정하였다.

DNN, CNN, 랜덤포레스트 모두 오버샘플링을 하지 않은 학습 데이터 실험도 함께하여 오버샘플링이 모형 학습에 미치는 효과를 확인하였다.

결과적으로 DNN, CNN, 랜덤포레스트 세 모형에 대해 오버샘플링하지 않은 학습과 오버샘플링 후의 학습을 진행하여 총 15개의 모델을 생성, 비교하였다.

모델 성능의 비교 지표는 Receiver Operating Characteristic(ROC) 곡선 아래의 면적인 Area Under Curve(AUC)와 이상거래 클래스에 대한 재현율(Recall) 값을 이용하였다. AUC는 불균형 데이터 세트에 대한 분류 모형의 성능을 확인할 때 가장 많이 사용되는 지표(Haixiang *et al.*, 2017)이며, AUC 값이 1에 가까울수록 모형의 성능이 우수한 것으로 볼 수 있다. 이상거래에 대한 재현율은 실제 이상거래 중 모형이 이상거래로 맞춘 비율로(식 (5)), 이상거래를 놓치지 않고 예측하는데 의의를 가지는 FDS 모형에서 주요 지표라고 볼 수 있다.

$$\text{이상거래 Recall} = \frac{True\ \ Positive}{True\ \ Positive + False\ \ Negative} = \frac{\text{이상거래를 이상거래로 맞춤}}{\text{실제 이상거래}} \quad (5)$$

<표 3>에 따르면, 모든 분류기에 대해 GAN 오버샘플링 기법을 사용한 것이 학습 성능을 높였다는 것을 알 수 있다. 특히 WGAN과 WCGAN의 경우Wasserstein Distance의 도입을 통해 위조데이터를 보다 실제와 유사하게 생성하여, 높은 학습 성능을 드러낸 것으로 생각된다. 그러나 cGAN의 경우 GAN과 비교하여 큰 차이를 보이지 않았는데, 이는 cGAN으로 위조데이터를 생성하는 과정에서 라벨의 역할이 크지 않았기 때문으로 생각된다. 이는 본 연구에서 사용된 데이터가 Kaggle 데이터이며 이상거래 내에 별다른 군집이 존재하지 않았기 때문으로 보인다. 그러나 현업에서 이상거래 데이터를 분석하게 된다면 이상거래의 패턴에 어떤 군집이 존재할 지 알 수 없으므로, cGAN의 적용이 필요할 것으로 사료된다. 또한, cGAN에 라벨을 부여하는 과정에서 탐색적 데이터 분석을 통해 이상거래 패턴의 군집 개수를 파악하고 K-평균 기법을 시행하거나, 밀도기반 군집 기법인 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)(Ester *et al.*, 1996) 등을 시도한다면, cGAN 모형 학습 과정에서 라벨의 유용성이 증가할 것으로 보인다. 또한 제안 모형 모두가 랜덤포레스트에 비해 높은 AUC 값을 가져 딥러닝 분류기가 높은 분류 성능을 가짐을 알 수 있다. 제안 모형 중 가장 높은 AUC값을 가진 모형은 DNN-WGAN 모형으로, 동일 오버샘플링 데이터에 대한 랜덤포레스트 모형에 비해 0.06 가량 향상되어 0.9857의 AUC값을 보였다. DNN 모형과 CNN 모형을 비교하였을 때, 오버샘플링 기법에 따라 DNN이 높은 AUC값을 가지기도 하나, 대체로 CNN 모형이 좋은 성능을 보였다.

〈표 3〉 오버샘플링 기법 별 각 분류기 성능 (AUC, Recall)

분류기	오버샘플링 기법	AUC	Fraud Recall
DNN	Original	0.959107	0.789474
	GAN	0.974032	0.789474
	cGAN	0.974378	0.769737
	WGAN	0.985697	0.789474
	WCGAN	**0.977982**	0.796053
CNN	Original	0.972080	0.789474
	GAN	0.975047	0.796053
	cGAN	0.978152	**0.802632**
	WGAN	0.978399	0.769737
	WCGAN	0.978733	0.796053
Random Forest	Original	0.910690	0.717105
	GAN	0.913723	0.769737
	cGAN	0.916118	0.736842
	WGAN	0.926209	0.756579
	WCGAN	0.916128	0.763158

재현율을 기준으로 모형을 비교한 경우에도 제안 모형의 성능이 랜덤포레스트에 비해 뛰어나 이상거래 탐지 목적에 부합하는 성능을 가졌음을 보였다.

5.2 한계점 및 제언

대부분의 금융기관 데이터의 경우 실시간 데이터 웨어하우스(Real-time Data Warehouse, RDW)에 적재된다. 따라서 금융기관의 분석 효율성을 높이기 위해선 실시간 분석과 분석 시간을 단축하는 프로세스가 필요하다. 그러나 본 논문에서는 기존에 적재된 데이터를 사용한 사례이기 때문에, 향후 차원 축소 기법 등을 도입하여 효율적인 데이터 수집/처리 방안에 대해 연구한다면 보다 산업 친화적인 FDS를 구축할 수 있을 것이라 생각한다. 또한 시간대/시퀀스를 고려할 수 있는 딥러닝 기법인 순환 신경망(Recurrent Neural Net, RNN)(Rumelhart *et al.*, 1987)을 분류기에 적용하면 특정 객체(사람 등)의 거래 패턴을 반영할 수 있고, 이상 패턴에 대한 해석력을 높일 수 있을 것으로 생각된다.

또한 본 연구의 실험 과정에서 사용된 데이터는 이미 주성분 분석이 완료된 Kaggle 데이터이므로, 실제 비즈니스 데이터에 적용할 경우 발생할 수 있는 불확실성이 존재한다. 그러나 이번 연구를 바탕으로 실제 비즈니스에서 사용하는 규칙 기반의 FDS 모형과 본 연구가 제안하는 다양한 딥러닝 기반의 FDS 모형을 하이브리드로 연결한다면FDS의 형태를 한층 진화시킬 수 있을 것이라 기대한다.

참 고 문 헌

[1] 김량형, 유동희, 김건우, “데이터마이닝 기법을 이용한 기업부실화 예측 모델 개발과 예측 성능 향상에 관한 연구”, *Information Systems Review*, 제18권, 제2호, 2016, pp. 173-198.

[2] 김주현, 원정임, “비지도학습 딥러닝을 활용한 이상거래탐지 시스템 모델”, *한국정보과학회 학술발표논문집*, 2018, pp. 917-919.

[3] 김한용, 이우주, “불균형적인 이항 자료 분석을 위한 샘플링 알고리즘들: 성능비교 및 주의점”, *Korean Journal of Applied Statistics*, 제30권, 제5호, 2017, pp. 681-690.

[4] 박재훈, 김휘강, 김은진, “의사결정나무를 이용한 이상금융거래 탐지 정규화 방법에 관한 연구”, *Journal of The Korea Institute of Information Security & Cryptology*, 제25권, 제1호, 2015, pp. 133-146.

[5] 서상현, 전용진, 이종수, 정호재, 김준태, “불균형 빅데이터의 효율적인 분류를 위한 생성적 적대 신경망 기반 오버샘플링 기법”, *한국정보과학회 학술발표논문집*, 2017, pp. 1030-1032.

[6] 손민재, 정승원, 황인준, “Conditional GAN을 활용한 오버샘플링 기법”, *한국정보처리학회 추계학술대회 논문집*, 제25권, 제2호, 2018, pp. 609-612.

[7] 이용현, 구해모, 김형주, “오토인코더를 활용한 효율적인 신용카드 사기 탐지 지도 기법”, *정보과학회컴퓨팅의 실제논문지*, 제25권, 제1호, 2019, pp. 1-8.

[8] Arjovsky, M., S. Chintala, and L. Bottou, “Wasserstein gan”, arXiv preprint, arXiv:1701.07875, 2017.

[9] Burez, J. and D. Van den Poel, “Handling class imbalance in customer churn prediction”, *Expert Systems with Applications*, Vol.36, 2009, pp. 4626-4636.

[10] Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique”, *Journal of Articial Intelligence Research*, Vol.16, 2002, pp. 321-357.

[11] Dal Pozzolo, A., O. Caelen, R. A. Johnson, and G. Bontempi, “Calibrating probability with under-sampling for unbalanced classification”, In Computational Intelligence, 2015 IEEE Symposium Series on, 2015, pp. 159-166.

[12] Douzas, G. and F. Bacao, “Effective data generation for imbalanced learning using conditional generative adversarial networks”, *Expert Systems*

with Applications, Vol.91, 2018, pp. 464-471.

[13] Ester, M., H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", In *Proceedings of Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.

[14] Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?", *Journal of Machine Learning Research*, Vol.15, No.1, 2014, pp. 3133-3181.

[15] Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative Adversarial Nets*, NIPS' 2014, 2014.

[16] Haixiang, G., L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data", *Review of Methods and Applications*, Vol.73, 2017, pp. 220-239.

[17] Ham, J., Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, No.3, 2005, pp. 492-501.

[18] Haykin, S., *Neural Networks and Learning Machines*, Pearson Prentice-Hall, New York, NY, 2009.

[19] He, H. and E. A. Garcia, "Learning from imbalanced data", *IEEE Transactions on Knowledge and Data Engineering*, Vol.21, No.9, 2009, pp. 1263-1284

[20] Hearst, M. A., S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines", *IEEE Intelligent Systems and Their Applications*, Vol.13, No.4, 1998, pp. 18-28.

[21] Hinton, G. E., S. Osindero, and Y. The, "A fast learning algorithm for deep belief nets", *Neural Computation*, Vol.18, 2006, pp. 1527-1554.

[22] Liaw, A. and M. Wiener, "Classification and regression by random Forest", *R News*, Vol.2, No.3, 2002, pp. 18-22.

[23] Mirza, M. and S. Osindero, "Conditional generative adversarial nets", arXiv preprint, arXiv:1411.1784, 2014.

[24] Murphy, K., *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[25] Purushu, P., N. Melcher, B. Bhagwat, and J. Woo, "Predictive analysis of financial fraud detection using azure and spark ML", *Asisa Pacific Journal of Information Systems*, Vol.28, No.4, 2018, pp. 308-319.

[26] Rumelhart, D. E., G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", *Parallel Distributed Processing*, Vol.1, 1987, pp. 318-362.

[27] Wang, J., J. Yang, S. L. Xiao, and D. Zhou, "Face recognition based on deep learning", *Human Centered Computing*, 2014, pp. 812-820.

[28] Zheng, P., S. Yuan, X. Wu, J. Li, and A. Lu, "One-class adversarial nets for fraud detection", arXiv preprint, arXiv:1803.01798, 2018.

Information Systems Review
Volume 22 Number 1
February 2020

Fraud Detection System Model Using Generative Adversarial Networks and Deep Learning

Ye Won Kim* · Ye Lim Yu** · Hong Yong Choi***

Abstract

Artificial Intelligence is establishing itself as a familiar tool from an intractable concept. In this trend, financial sector is also looking to improve the problem of existing system which includes Fraud Detection System (FDS). It is being difficult to detect sophisticated cyber financial fraud using original rule-based FDS. This is because diversification of payment environment and increasing number of electronic financial transactions has been emerged. In order to overcome present FDS, this paper suggests 3 types of artificial intelligence models, Generative Adversarial Network (GAN), Deep Neural Network (DNN), and Convolutional Neural Network (CNN).

GAN proves how data imbalance problem can be developed while DNN and CNN show how abnormal financial trading patterns can be precisely detected. In conclusion, among the experiments on this paper, WGAN has the highest improvement effects on data imbalance problem. DNN model reflects more effects on fraud classification comparatively.

Keywords: Fraud Detection System, Deep Neural Net, Convolutional Neural Net, Generative Adversarial Network

* Finance/Public Big Data Team Data Science Associate, LG CNS
** Finance/Public Big Data Team Data Science Associate, LG CNS
*** Corresponding Author, Finance/Public Big Data Team Data Science Leader, LG CNS

◐ 저 자 소 개 ◑

김 예 원 (yewon.kim@lgcns.com)

고려대학교 통계학과에서 학사를 마쳤으며 2018년부터 LG CNS에 재직 중이다. 현재 국세청 빅데이터 플랫폼 구축사업에 참여하고 있다. 주요 연구 관심분야는 머신러닝, 딥러닝, 강화학습이다.

유 예 림 (yelim.yu@lgcns.com)

홍콩과학기술대학교 Maths-Statistics and Financial Maths에서 학사를 마쳤으며 2018년부터 LG CNS에 재직 중이다. 현재 국세청 빅데이터 플랫폼 구축사업에 참여하고 있다. 주요 연구 관심분야는 머신러닝, 딥러닝, 강화학습이다.

최 홍 용 (hychoi@lgcns.com)

부산대학교 물리학과에서 학사를 마쳤으며, 동 대학원에서 경영학으로 석사학위를 취득하였다. 현재 LG CNS에 재직 중이다. 국세청, LG U+, LG전자, LG화학, 범한판토스, LS산전 등 다수의 빅데이터 및 Business Intelligence 프로젝트에 참여하였다. 주요 연구 관심분야는 빅데이터, Business Intelligence, EPM 등이 있으며, 현재 금융/공공 빅데이터팀장을 담당하고 있다.

논문접수일 : 2019년 07월 18일 **게재확정일 : 2019년 10월 07일**
1차 수정일 : 2019년 09월 25일