



금융업에서의 딥러닝 적용 사례와 모델 개발 과정

📅 속성	@2021년 1월 12일
🏷️ 태그	논문 이상거래탐지 자산분배

📖 서론

1. 금융업내 딥러닝의 적용 사례

- 1-1. 맞춤형 금융 챗봇(Chatbot)
- 1-2. 맞춤형 포트폴리오 서비스 제공
- 1-3. 자금세탁방지과 이상거래탐지 시스템 구축
- 1-4. 금융회사 준법감시시스템의 활용
- 1-5. 금융업 딥러닝 도입 동향 및 이슈

2. 딥러닝을 활용한 자산분배 시스템 개발 과정

- 2-1. 시스템 구성
- 2-2. 실험 방법
- 2-3. 결과분석

3. 딥러닝을 이용한 전자금융이상거래 탐지 모델 수립

- 3-1. 연구 배경
- 3-2. 연구 방법
- 3-3. 1차 실험
- 3-4. 2차 실험
- 3-5. 실험의 결론

📖 결론

cf) 개인 의견 🧑

📁 참고문헌

📖 서론

- 금융시장 : 금리 ↓ → 예금, 적금 보다는 금융투자 등의 공격적인 재테크 전략에 대한 관심 ↑
- 금융업계 : 높은 리스크로 인해 투자를 주저하는 소비자를 공략하기 위해 딥러닝 기법을 활용한 금융상품 및 서비스 개발 ↑
- 딥러닝 기술 발달 → 리스크 감소, 비용 절감, 생산성 증대, 고객맞춤 서비스 강화 등의 효과 기대
- 2016년 인공지능 시스템에 가장 많은 투자를 한 산업 ▶▶ 금융, 소매 산업
- 앞으로 인공지능 기술 활용도가 높을 것으로 기대되는 분야 ▶▶ 의료, 금융, 스마트 홈, 교통, 제조 순



금융산업에서 인공지능 기술이 활용되는 분야

1. 맞춤형 금융상담과 콜서비스
2. 주식과 채권 등 자산관리와 효과적인 재산배분
3. 차세대 자금세탁방지과 이상거래 탐지
4. 준법감시시스템 - 레그테크(RegTech) 도입
5. 의사결정을 지원하는 분석 보고서 작성

1. 금융업내 딥러닝의 적용 사례

1-1. 맞춤형 금융 챗봇(Chatbot)

챗봇을 도입해 금융상담, 콜서비스와 같은 고객대응 업무에 사용

채팅 로봇인 챗봇은 모바일 기기에서 인공지능 채팅앱의 형태를 통해 고객에게 맞춤화된 서비스를 제공하는 기능

- 신한카드 : 고객이 설정한 소비 항목에 맞추어 소비 내역을 분류하여, 고객에게 적절한 소비 항목을 추천
- 삼성카드 : 고객의 결제 정보 및 주변 상권 정보를 학습하여 고객 동선과 점포 이용 성향 분석

1-2. 맞춤형 포트폴리오 서비스 제공

로보어드바이저를 이용한 주식, 채권, 외환 등 자산관리 / 효과적인 자산배분 가능


주가 예측 정보 제공 및 고객의 투자성향과 투자금액에 따라 맞춤형 포트폴리오 서비스를 제공

- KEB하나은행 : 인공지능 기반의 로보어드바이저 'HAI Robo' 출시

(일반적으로 제공되는 과거 수익률, 변동성 외에도 자산 분산도, 비용 효율성, 시뮬레이션 등 다양한 정보 제공)

1-3. 자금세탁방지과 이상거래탐지 시스템 구축

자금세탁 등 범죄가 의심스러운 거래, 2천만원 이상의 고액현금거래, 외국환거래정보 등을 탐지

인공지능 기술로 새로운 자금세탁 패턴이나 유형을 학습하고  적발할 수 있도록 구축할 방침

- 한국거래소 : 시세조종 혐의 계좌와 정상 계좌의 데이터를 학습하고, 시세조종 개연성을 판단하는 인공지능경망 기반의 감시모델 개발
- SK증권 : 기존의 룰 기반과 더불어 딥러닝 기술을 적용해 스스로 데이터를 학습하고, 금융사고의 패턴을 분리해 금융사고를 예방하는 최신의 금융사기탐지 방식을 병행 적용할 계획

1-4. 금융회사 준법감시시스템의 활용

금융규제 및 컴플라이언스 요구사항을 효과적으로 해결하고자 금융회사들은 레그테크(Reg Tech) 도입

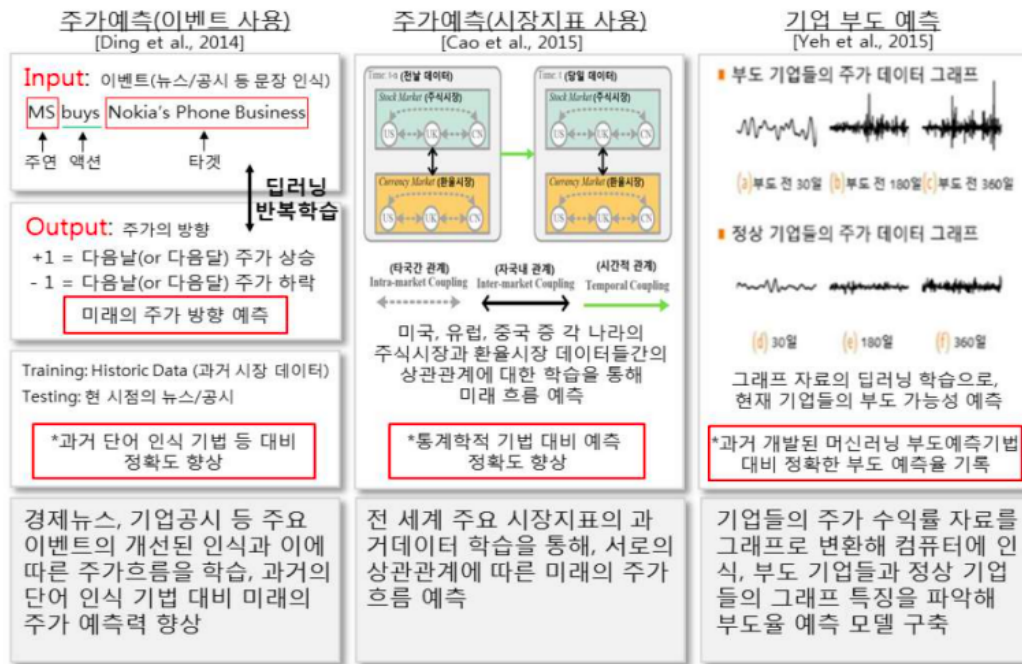
레그테크(Reg Tech) ? 규제를 뜻하는 레귤레이션(Regulation) + 기술을 뜻하는 테크놀로지 (Technology)

→ 금융회사로 하여금 내부통제와 법규준수를 용이하게 하는 정보기술을 의미

활용분야: 데이터 관리, 위험분석 및 예측, 금융거래 탐지, 고객신원관리, 기업내부통제, 컴플라이언스 등

1-5. 금융업 딥러닝 도입 동향 및 이슈

- 세계 1위 투자은행인 골드만삭스는 주식 트레이딩에 인공지능 켄쇼(Kensho)를 도입해 600여명에 달했던 주식 매매 트레이더들을 현재 2명까지 줄임
- 투자자문 및 트레이딩 분야: 딥러닝 기술을 사용하여 현재 경제와 금융 시장의 상황을 분석하고 미래를 예측



(그림 44) 딥러닝 기술을 이용한 금융시장 예측 알고리즘

(출처: 김희민, “알파고의 딥러닝 금융업 적용 사례”, KB금융지주경영연구소(2016.4.25))

2. 딥러닝을 활용한 자산분배 시스템 개발 과정

딥러닝을 활용한 자산분배 시스템

딥러닝 네트워크 기반의 알고리즘의 발전으로 인공지능은 전례적으로 빠른 성장세를 보이고 있다. 그 중 금융은 인공지능이 가장 많이 활용될 분야로 예상되고 있으며 최근 많은 연구가 되고 있다. 기존의 딥러닝을 사용한 재무 전략은 단일 종목에 대한 주가 예측에만 치중되어 있어 변동성에 취약하다. 따라서 본 연구는 딥러닝을 이용하여 펀드

○ <https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201912937050956&SITE=CLICK>

SCIENCE ON

김성수, 김종인, 정기철(2019), 「딥러닝을 활용한 자산분배 시스템 = Portfolio System Using Deep Learning」, 한국산업정보학회논문지 = Journal of the Korea Industrial Information Systems Research v.24 no.1(pp.23 - 30), 한국산업정보학회

본 논문에서는 인공 신경망을 활용한 ETF 상품을 만드는 포트폴리오 시스템을 제안

→ 딥러닝 신경망을 통해 코스피 100지수의 종목들 중 일부 종목을 추출하고, 추출된 종목들을 ETF 상품을 만드는 시스템을 제안
(코스피 100지수 : 시가총액이 큰 상위 100종목)



ETF(상장지수펀드) ? KOSPI, KOSDAQ과 같은 특정 지수의 성과를 추적하는 인덱스 펀드를 주식처럼 거래할 수 있게 한 펀드

📌 예시 : KODEX 200이라는 ETF는 KOSPI 200 지수의 약 100배 가격으로 책정, 즉 KOSPI 200 지수가 증가하면 KODEX 200도 증가, 감소하면 같이 감소

+) 펀드 → 1주만 보유해도 ETF 상품을 구성하는 모든 종목의 성과를 누릴 수 있는 장점

2-1. 시스템 구성

본 시스템의 주 데이터인 주가 데이터는 시계열 데이터 (특정 시간에 측정된 데이터)를 활용하기 때문에 시계열 처리에 적합한 신경망 구조가 필요 → 오토인코더 신경망 구조를 사용



오토인코더 : 딥러닝의 대표적인 비지도 학습 알고리즘, 데이터의 잡음 제거와 패턴 추론에 강점을 지님

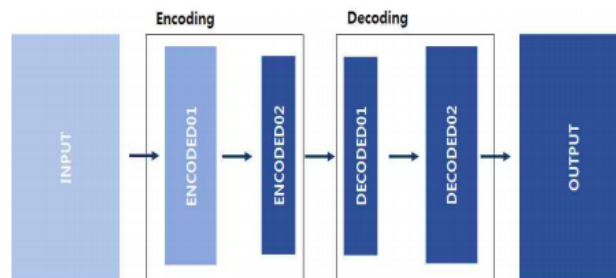


Fig. 1 Autoencoder Model Structure

→ 오토인코더 신경망은 Fig. 1과 같이 신경망층 두 개를 대칭되게 붙여 놓은 구조

입력층을 통해 들어온 데이터는 인코딩층을 거쳐 특징 값을 구하고, 디코딩층에서는 특징 값을 복원하여 최종 출력 층에서 원본과 최대한 유사하도록 학습 하는 신경망 구조

→ 과거의 주가 데이터의 움직임을 학습해 미래의 주가의 추세를 예측하는데 적합한 신경망 구조

본 연구에서 사용하는 데이터는 2014년 1월 2일에 코스피 100 지수에 구성된 종목들의 일별 주가 데이터이다.

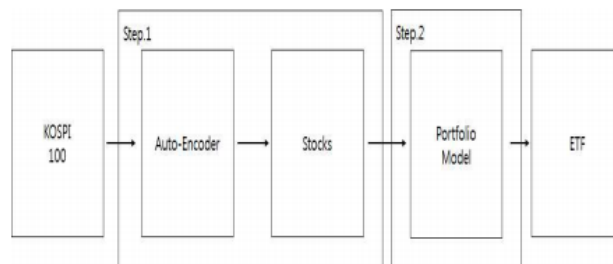


Fig. 2 Deep Learning Portfolio Process

본 논문은 위와 같이 설계된 모델을 사용

모든 신경망층은 과적합을 방지하기 위해 L2 정규화를 사용

800epoch 동안 학습을 진행

오토인코더 신경망의 인코딩층과 디코딩층의 활성화 함수로는 ReLU를 사용

출력층의 활성화 함수는 선형 함수를 사용

손실함수로는 Mean Squared Error를 사용

최적화 함수로는 Adam Optimizer 함수를 사용

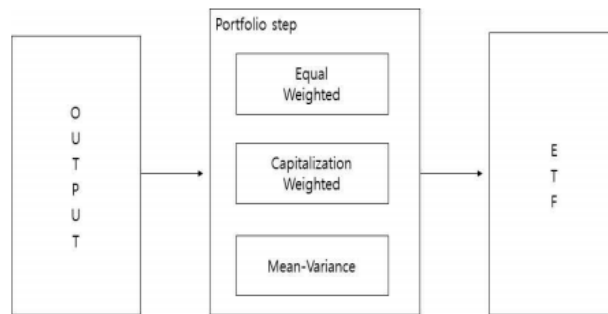


Fig. 3 Portfolio Selection Model



오토인코딩 구간에서 선택된 종목들을 아래의 3가지 방법으로 하나의 ETF 상품으로 만듦

- 동일비중 포트폴리오(Equal-weighted) : SNP500 지수를 만드는 방법

위험자산 N개에 대해 각 종목의 비중 W를 1/N로 동일하게 구성

$$W_x = \frac{1}{N}$$

- 시가총액 가중 포트폴리오(CapitalizationWeighted) : KOSPI 지수를 만드는 방법

종목 x의 구매비중 W를 위험자산 x의 시가총액 / 구성 종목들의 시가총액 총합

$$W_x = \frac{C_x}{\sum C_x}$$

C_x : 종목 x의 시가총액

시가총액 : 주가와 발행 주식수를 곱한 것, 기업 가치를 평가하는 지표

(순위 : 삼성전자, SK하이닉스, LG화학 ...)

- 평균-분산 포트폴리오(Mean-Variance)

N개의 위험자산에 대하여 각 종목들의 평균과 분산을 사용해 최적의 포트폴리오를 찾는 방법

$$W_x = \text{MAX}\left(\frac{\text{mean}(R_x) - R_f}{\sqrt{\text{VAR}}}\right)$$

R_x : 종목 x 수익률, R_f : 무위험 수익률, VAR : 종목분산

2.2. 실험 방법

데이터는 첫 800 거래일을 기준으로 training data와 test data로 나누었으며,

training data중 마지막 150일은 실험 모델의 과적합을 방지하기 위해 validation data로 사용함

test data는 2017년 4월 1일부터 2018년 4월 17일까지이며, 이 시점에 계산된 구매비율로 구성된 펀드의 성능을 평가
시계열 데이터인 주가 데이터를 사용하기 때문에 데이터를 섞지 않았음



모델 평가할 때 사용할 결정 변수

- 수익률 : 포트폴리오가 얼마만큼 수익을 더 보였는가에 대한 지표

$$Mean = \sqrt[n]{\prod_i price_{x,i}}$$

$price_{x,i}$: i 시점의 종목 x 의 가격

- 샤프지수 : 위험 자산에 투자함으로써 얻은 초과 수익의 정도를 나타내는 지표
→ 지수가 높을수록 수익을 변동 폭이 크지 않으면서 높은 수익을 올리는 펀드

$$SharpeRatio = \frac{ER_x - ER_f}{\sigma_x}$$

σ_x : 자산의 표준편차, ER_x : 자산 수익률, ER_f : 무위험수익률

- 최댓값 대비 하락률 : 투자 기간 중 포트폴리오의 고점에서 저점까지 최대 누적손실
→ 만들어진 펀드가 하락장에서 얼마나 가격 방어를 할 수 있는지를 나타내는 지표

$$MDD = MAX(\frac{Price_{x,i} - Peakprice_{x,i}}{Peakprice_{x,i}})$$

$Price_{x,i}$: i 시점 x 의 가격, $Peakprice_{x,i}$: i 시점 x 의 최고가격

2-3. 결과분석

Table 1.1 Result of Experiment(30 stocks)

Funds	Average Return	Sharpe ratio	MDD
KOSPI100	11.48	21.07	-11.01
Equal-Weighted	8.82	26.89	-10.96
Capitalization-Weighted	9.78	22.27	-10.80
Mean-Variance	17.75	10.56	-14.30

Table 1.2 Result of Experiment(50 stocks)

Funds	Average Return	Sharpe ratio	MDD
KOSPI100	11.48	21.07	-11.01
Equal-Weighted	8.79	29.07	-10.06
Capitalization-Weighted	8.52	24.91	-9.83
Mean-Variance	17.77	9.97	-19.8

- 평균-분산 포트폴리오를 통해 구성된 펀드: 가장 높은 수익률, 가장 낮은 샤프비율, 가장 높은 최대값 대비 하락률 → 변동성이 크다는 단점
- 동일비중 포트폴리오 및 시가총액 가중 포트폴리오: 코스피 100 지수보다 더 높은 샤프비율, 더 낮은 최대값 대비 하락률 → 변동성이 더 낮다는 장점 **BUT!** 코스피 100 지수에 비해 더 낮은 수익률

따라서 본 논문에서 제안하는 시스템을 사용하여 포트폴리오를 만들 경우 **더 높은 수익률** 혹은 **더 적은 변동성** 을 지닌 펀드를 만들 수 있음

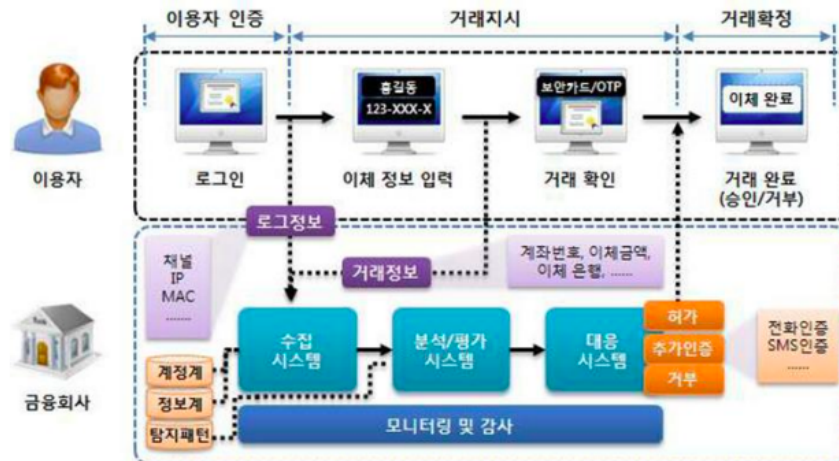
3. 딥러닝을 이용한 전자금융이상거래 탐지 모델 수립

http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=7c4a4562c113032affe0bdc3ef48d419

전일교(2017), 「딥러닝을 이용한 전자금융이상거래 탐지 모델 수립」, 서울 : 연세대학교 정보대학원

3-1. 연구 배경

[그림1]FDS 시스템 개요³⁾



→ 국내 은행에서 운영중인 FDS 시스템 (전자금융이상거래탐지시스템)

1. 비대면 채널에서 이루어지는 거래 로그를 수집하고 분석
2. 이상거래/정상거래 해당 여부를 판단
3. 이상거래로 의심될 경우 추가인증(ARS, SMS) 또는 거래제한

하지만, 노하우가 부족하고 시나리오 기반의 룰탐지 수준에 머무르고 있음

→ **90% 이상의 오탐율**을 보여 고객의 불편과 정책 운영의 어려움을 초래

3-2. 연구 방법

본 논문에서는 1차와 2차로 연구를 나누어 진행

- 1차 연구: 2016년도 전자금융 피해사고가 발생한 전자금융 이용자의 2년치 데이터를 수집하여 이용자별 프로파일링 분석을 통해 데이터를 전처리한 후 학습의 input data로 활용, output data로 정상거래/사고거래 구분 값을 입력하여 DNN을 통한 모델을 생성 및 검증
- 2차 연구: 예상되는 미래의 전자금융환경에 맞게 1차 실험에 사용되었던 학습데이터를 가공하여 1차와 마찬가지로 DNN을 통한 모델을 생성 및 검증을 수행

3-3. 1차 실험

2016년도 전자금융사고 피해고객 113명의 과거 2년치 거래 Raw데이터를 수집한 후,

매 이체거래 마다 과거 1년치 거래에 대한

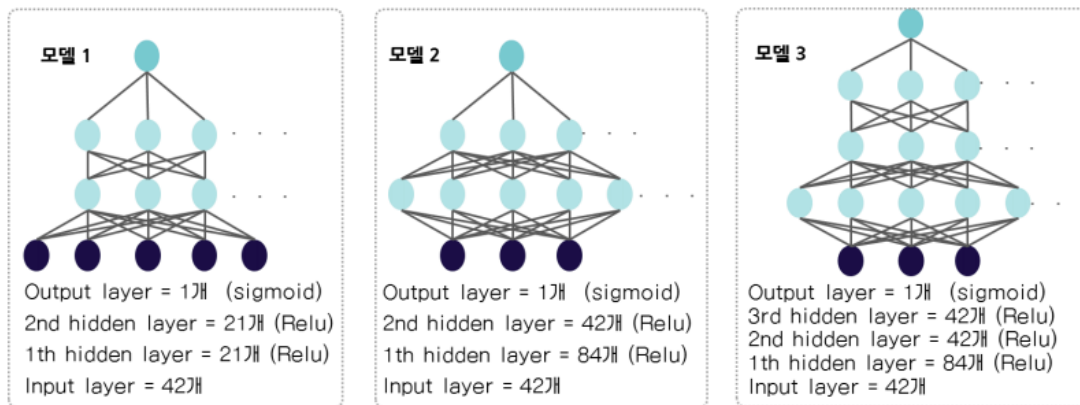
사용자의 전자금융 이용환경(OS환경, 네트워크환경),

거래성향(거래시각, 거래패턴 등)에 대해 프로파일링 분석을 통해

Raw 데이터를 Featured 데이터로 전처리해 input data로 사용

DNN의 설계는 3가지 서로 다른 형태의 Topology로 설계 → 은닉층의 깊이는 3depth를 넘지 않도록 설계

[그림3]전자금융이상거래탐지 모델 Topology



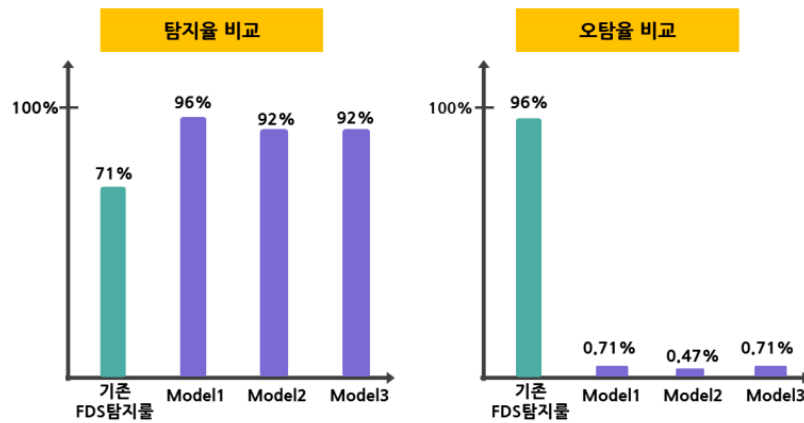
모델 설계

- 은닉층에서의 활성화 함수는 ReLU를 적용
- 출력층에서는 sigmoid를 적용 → 사고거래(1), 정상거래(0)을 학습하도록 설계
- 손실함수로는 crossentropy 적용
- 최적화 알고리즘으로는 Adam을 적용

[그림5]전자금융이상거래탐지 모델 검증결과

모델 1 검증결과	모델 2 검증결과	모델 3 검증결과
학습횟수 : 40회 batch_size : 10 사고탐지율 : 27건 중 26건 탐지 (96%) 사고오탐지율 : 1265건 9건 오탐 (0.71%)	학습횟수 : 40회 batch_size : 10 사고탐지율 : 27건 중 25건 탐지 (92%) 사고오탐지율 : 1265건 6건 오탐 (0.47%)	학습횟수 : 40회 batch_size : 10 사고탐지율 : 27건 중 25건 탐지 (92%) 사고오탐지율 : 1265건 9건 오탐 (0.71%)
Total fraud count : 27 Detect fraud count : 26 Percent of Detect fraud from total fraud : 96.00% Total normal count : 1265 Detect fraud count : 9 Percent of Detect fraud from total normal : 0.71%	Total fraud count : 27 Detect fraud count : 25 Percent of Detect fraud from total fraud : 92.00% Total normal count : 1265 Detect fraud count : 6 Percent of Detect fraud from total normal : 0.47%	Total fraud count : 27 Detect fraud count : 25 Percent of Detect fraud from total fraud : 92.00% Total normal count : 1265 Detect fraud count : 9 Percent of Detect fraud from total normal : 0.71%

[그림6] 기존FDS탐지율과의 비교



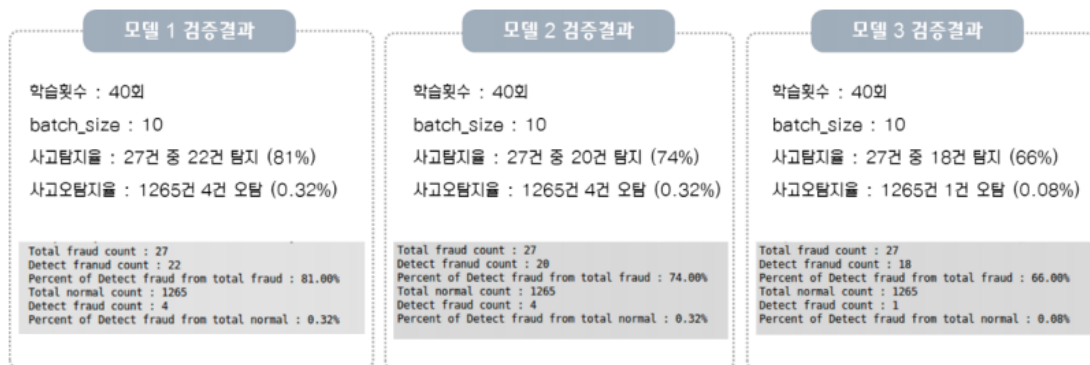
→ 모든 모델이 기존의 FDS 시스템보다 탐지율이 높고, 오탐율이 확연히 낮음을 알 수 있음

3-4. 2차 실험

최근 금융감독당국에서는 인터넷뱅킹 서비스에 **보안프로그램 설치 의무화를 폐지함**

→ 인터넷뱅킹서비스 사용자의 OS식별정보, 네트워크 설정 정보, 브라우저정보를 수집할 수 없으므로 이와 관련한 학습데이터 **변수 10개를 제외한 32개의 변수**를 학습데이터의 input data로 재정의

[그림7] 전자금융이상거래탐지 모델 검증결과(2차 실험)



→ 67%~81%로 1차 실험에 비해 낮은 사고거래 탐지율, 오탐율의 경우 0.5% 미만의 비교적 낮은 오탐율

→ 학습데이터에 사용되는 **변수의 수가 축소**됨에 따라 정상거래와 사고거래를 분류하는 특징의 정확도 ↓

3-5. 실험의 결론

딥러닝을 적용한 전자금융 이상거래탐지 모델 → 기존 룰기반 대비 높은 탐지율과 현격하게 낮은 오탐율을 달성

→ 딥러닝 기술이 국내 **전자금융 이용자 보호**를 위해 유의미하게 활용될 수 있음

결론

http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=4eec2e5b63bcfc59ffe0bdc3ef48d419

박지훈(2020), 「딥러닝 기반의 금융회사 고객이탈 예측모형에 관한 연구」, 전주: 전북대학교 일반대학원

- 2019년 금융위원회 → 디지털 전자금융 규제에 새로운 패러다임을 제시 → 금융기관간의 경쟁력 더욱 심화
- 2019년 7월 **오픈뱅킹** 법제도와, 마이페이먼트(지급지시전달업) 도입 등 → 전자금융 산업체계와 진입규제, 영업행위 규제를 보다 현대화할 계획
- 하지만 사회 과학 서비스 분야 중 특히 **금융시장** 부문에서 딥러닝 기술은 아직 **기초적인** 단계에 머무르고 있음 → **금융관련 데이터** 가 가지고 있는 개별적인 속성 자체(**정보의 비대칭과 데이터 희소성**)와 분석 결과에 대한 **해석의 어려움** 때문
- 하지만 더 나은 기술 개발을 위해 데이터 관련 제도가 개선되고 있기 때문에, 기술 발전에 빠르게 대응해야함

cf) 개인 의견

딥러닝에 대한 공부 없이 논문 스터디를 했다면, sigmoid, ReLU, crossentropy 등의 개념처럼 이해가 되지 않았을 부분도 많았을 것입니다. 논문을 읽으면서 아는 부분이 나올 때는 반갑기도 하고, 공부한 보람을 느끼기도 했습니다. 그리고 그동안 배운 개념이 제가 관심있는 분야에서 어떻게 쓰이는지 알게 되면서, 딥러닝 기술 중에서도 어느 부분을 더욱 중점적으로 공부할 것인지, 추후에 관련 분야에서 일할 때, 어떻게 적용할지에 대한 생각을 할 수 있어서 좋았습니다.

참고문헌

- 박원기(2018), 「딥러닝 소개와 금융업 적용 사례」, 서울 : 고려대학교 정책대학원
- 김성수, 김종인, 정기철(2019), 「딥러닝을 활용한 자산분배 시스템 = Portfolio System Using Deep Learning」, 한국산업정보학회논문지 = Journal of the Korea Industrial Information Systems Research v.24 no.1(pp.23 - 30), 한국산업정보학회
- 전일교(2017), 「딥러닝을 이용한 전자금융이상거래 탐지 모델 수립」, 서울 : 연세대학교 정보대학원
- 박지훈(2020), 「딥러닝 기반의 금융회사 고객이탈 예측모형에 관한 연구」, 전주: 전북대학교 일반대학원