정 보 보 호 학 회 지
제26권 제2호, 2016. 4

인공지능과 핀테크 보안

최 대 선*

요 약

본 논문에서는 핀테크 보안에 활용 가능한 딥러닝 기술을 살펴본다. 먼저 인공지능과 관련된 보안 이슈를 인공지능이 사람을 위협하는 상황에 대한 보안(Security FROM AI), 인공지능 시스템이나 서비스를 악의적인 공격으로부터 보호하는 이슈(Security OF AI), 인공지능 기술을 활용해 보안 문제를 해결하는 것(Security BY AI) 3가지로 구분하여 살펴본다. Security BY AI의 일환으로 딥러닝에 기반한 비정상탐지(anomaly detection)과 회귀분석(regression)기법을 설명하고, 이상거래탐지, 바이오인증, 피싱, 파밍 탐지, 본인확인, 명의도용탐지, 거래 상대방 신뢰도 분석 등 핀테크 보안 문제에 활용할 수 있는 방안을 살펴본다.

Ⅰ. 서 론

알파고와 이세돌의 바둑대결로 인공지능에 대한 관심이 고조되고 있으며, 각 분야에서 인공지능 적용이 활발하다. 핀테크 서비스는 간편결제를 넘어서 비대면 본인확인을 기본으로 하는 인터넷 전문은행, 빅데이터 분석을 통한 신용 예측 기반 대출서비스, 거래 상대방의 신뢰도 분석이 필요한 P2P대출, 크라우드펀딩 등으로 고도화되고 있다. 본 논문에서는 핀테크 보안에 활용할 수 있는 인공지능기술을 살펴본다.

먼저 인공지능과 관련된 보안 이슈를 분석하고, 인공지능 기술을 활용한 보안 솔루션 들을 살펴본다. 최신 인공지능 기술로 각광을 받고있는 딥러닝 기술 중에서 보안 문제에 적용 가능한 기술을 설명하고, 핀테크 보안에 적용하는 방안에 대해 살펴본다.

Ⅱ. 인공지능과 보안

인공 지능의 보안 이슈는 인공지능이 사람을 위협하는 상황에 대한 보안(Security FROM AI), 인공지능 시스템이나 서비스를 악의적인 사람들로부터 보안하는 이슈(Security OF AI), 인공지능 기술을 활용해 다른 보안 이슈를 해결하는 것(Security BY AI) 3가지로 구분할 수 있다.

2.1. Security FROM AI

사람들은 인공지능과 바둑대결에서 이세돌9단이 계속 패하자 터미네이터를 연상하며 두려움에 빠졌다. 그리고 인공지능의 위험성에 대해 생각하기 시작했다. 많은 전문가들이 인공지능의 위험성에 대해 경고했는데 스티븐호킹은 “인공지능은 인간의 지능을 넘으면 스스로 더 나은 지능을 점점 더 빠른 속도로 설계할 수 있을 것이며, 인간은 생물학적 진화 속도가 느리기 때문에 경쟁에 뒤지고 인공지능에 의해 대체될 것이다”는 우려를 밝혔으며, 테슬라 CEO 엘런 머스크는 “인공지능은 인류 생존의 가장 큰 위협이다”라고 주장했다. 이러한 우려는 인공지능이 사람의 두뇌를 추월하는 강 인공지능 단계에 이르거나, 인간을 훨씬 초월하는 초 인공지능 단계에 이르렀을 때 발생 가능한 향후의 문제라고 할 수 있다.

한편, 인공지능의 윤리 문제는 이미 코 앞에 다가와 있다. 예를 들어, 무인 자동차가 도로를 달리던 중 무단횡단을 하는 보행자를 발견했다. 이때 자동차가 보행자를 피하면 탑승자가 가드레일과 충돌해 사망할 것이다. 그렇지 않으면 보행자가 사망하고 탑승자가 생존할 수 있는 상황이다. 이처럼 윤리적 선택이 필요한 상황에서 인공지능이 어떤 선택을 내리도록 만드는 것이 옳은가? 또한 무인 자동차의 과실로 인한 사고 책임은 누가 져

* 공주대학교 의료정보학과(sunchoi@kongju.ac.kr)

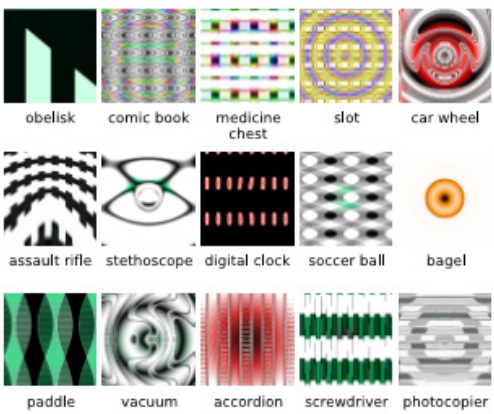
야 하는지, 생사에 관한 문제를 전적으로 인공지능에게 맡기는 것이 옳은지에 대한 판단이 필요하다[1].

2.2. Security OF AI

현재의 인공지능기술은 전통적인 컴퓨터 시스템 상에서 구현되고 있다. 따라서 인공지능 시스템을 악의적인 공격으로부터 보안하는 문제는 기존의 정보보안 이슈와 다르지 않다. 다만 인공지능 기술의 가까운 적용 대상으로 자주 언급되는 자율 주행 자동차에 대한 해킹이나 의료진단/수술 시스템에 대한 공격, 무인기를 비롯한 군사로봇에 대한 해킹 등의 문제는 심각한 피해를 야기할 수 있기 때문에 더욱 주목을 해야 한다.

기계학습을 기반으로 한 인공지능 시스템에 대한 새로운 유형의 공격 방법이 이슈가 될 수 있다. 마이크로소프트의 인공지능 채팅로봇 테이는 가동 하루만에 서비스를 중단했다. “유대인이 싫다” 같은 인종차별적 발언을 했기 때문인데, 이렇게 된 원인은 테이가 사람들과의 대화를 통해 대화패턴을 학습하고 말을 하는 시스템인데, 악의적인 의도를 가진 사람들이 서비스에 접속하여 인종차별적 발언을 지속적으로 학습시켰기 때문이다.

인공지능을 속일 수 있는 특이한 입력 패턴을 알아내면 인공지능의 동작을 왜곡시킬 수 있다. 그림1은 딥러닝 영상인식의 오류 사례이다. 사람이 보면 전혀 비슷하지 않지만, 영상인식에서는 잘못 인식된 패턴이다. 그런데, 모든 패턴을 사전에 시험할 수 없기 때문에 실제 오류가 발생하기 이전에는 이런 오류 패턴을 사전에 알기 어렵다. 이러한 오류 패턴은 소프트웨어에 포함된 취약점(security hole)과 마찬가지로 문제가 될 수 있다. 예를 들어 자율 주행자동차가 길로 인식할 수 있는 사진을 막다른 길에 붙여 놓으면 자율 주행자동차는 정지하지 않고 사고를 일으키게 될 것이다.



〔그림 1〕 딥러닝 영상인식 분류 오류〔2〕

이처럼 학습 데이터나 입력 데이터를 조작하여 기계학습 기반의 인공지능에 오동작을 유발할 수 있는데 이러한 Security OF AI 이슈에 대한 연구는 미흡한 실정이다.

2.3. Security BY AI

인공지능 기술, 특히 기계학습을 이용하여 보안 문제를 해결하는 연구는 오래 전부터 진행 되어왔다. 바이오 인증, 침입탐지, DDOS 탐지, 이상거래탐지 등 다양한 분야에서 머신러닝 기법이 사용되어왔다. 최근 각광받는 딥러닝을 이용한 보안 기술도 보고되고 있는데, 국내 사례는 표 1과 같다.

한편, 인공지능 기술을 이용한 해킹 위협이슈가 될 수 있다. 인간이 알아 낼 수 없는 보안 취약점을 인공지능의 탐색 기술이나 분석 기술을 이용하여 알아 낸 뒤, 이를 이용해 해킹을 하는 사례도 보고되고 있다[3].

〔표 1〕 딥러닝 기반 보안기술

분야	조직명	내용
바이오 인증	슈프리마	딥러닝기반 지문인증
	공주대학교	딥러닝기반 동적서명인식
네트워크 보안	한국전자 인증	인공지능기반 APT 공격 역추적
	유넷시스템	UBA와 머신러닝기반 통합로그분석
신용분석	솔리드웨어	딥러닝 기반 신용분석

Ⅲ. 인공지능기반 핀테크 보안 기술

핀테크 보안에 사용될 수 있는 인공지능 기술 중에서 딥러닝 기술이 높은 성능을 발휘할 수 있는 분야로 비정상탐지와 회귀분석이 있다.

3.1. 딥러닝 기반의 비정상 탐지 (Anomaly Detection)

비정상 탐지(Anomaly detection)는 기대되는 패턴에 부합되지 않거나 데이터셋 내의 타 아이템과 다른 아이

템을 찾아내는 과정을 의미한다[4]. 보안에 있어서 비정상패턴은 공격이나, 사기, 악성코드를 의미한다.

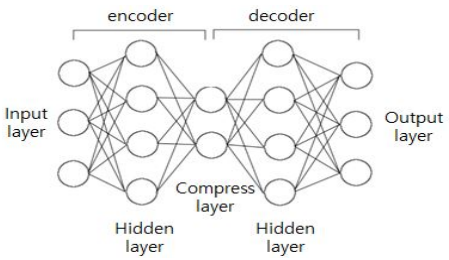
이러한 비정상 패턴을 탐지하는 방법으로 기존에는 시그니쳐나 규칙기반 탐지가 사용되어 왔으나, 최근에는 머신러닝을 이용하려는 시도가 많이 이루어지고 있다. 머신러닝을 이용한 비정상 탐지 방법은 1) 정상 패턴과 비정상 패턴을 학습하여 분류하는 지도학습 2) 클러스터링으로 패턴을 2가지로 분류하는 비지도학습 3) 정상패턴만을 학습하는 세미지도학습 방법이 있다.

그런데 보안 문제에 있어서는 비정상패턴의 수가 적으며, 새로운 유형의 공격, 사기, 악성코드 등이 등장하기 때문에 비정상패턴을 학습하는 방법으로는 새로운 유형의 공격에 대응할 수 없다. 클러스트링도 비정상패턴의 수가 적은 경우는 적용하기 어렵다. 따라서 정상패턴만을 학습하여 비정상패턴을 구분하는 방법을 사용해야한다. 이러한 방법으로는 1 class SVM [5]이 있는데, 높은 성능을 보이지는 못하고 있다.

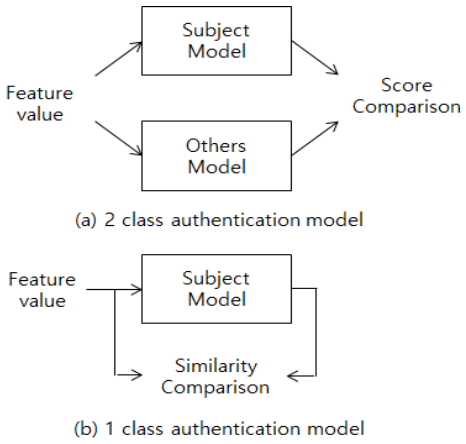
최근 각광을 받는 딥러닝 기술 중에는 1 class 분류기를 구성하는 네트워크는 없었다. 본 논문에서는 Auto-encoder를 사용한 비정상탐지 방법을 제안한다. Auto-encoder는 입력값에 따라 출력값을 재생산하는 신경망이다[6].

그림 2에 보이는 것처럼 Auto-encoder는 encoder과 decoder를 붙여 놓은 형태이며 중간에 compress layer로 연결된다. 학습(train) 단계에서는 정상데이터 만을 학습하게 되고, 예측(predict) 단계에서는 주어진 입력값에 대해 출력값을 생성하게 된다. 이때 학습데이터와의 유사도에 따라 값의 재현도가 달라진다.

즉 학습된 데이터와 유사한 패턴의 입력값일수록 출력값의 재현도도 높아지게 된다. 따라서 입력값과 출력값의 차이를 이용하여 비정상패턴을 탐지할 수 있다. Auto-encoder를 이용하여 모창가수를 구분하는 문제에 적용하여 사람과 동등한 성능을 얻었으며[7], 이러한 구



〔그림 2〕 Auto-encoder 구조



〔그림 3〕 authentication model

조에 기반한 이상거래탐지기술을 연구 중에 있다.

비정상탐지 기법은 핀테크 보안의 중요 요소인 인증과 본인확인, 명의도용탐지에도 적용될 수 있다. 특히 최근 각광을 받고 있는 바이오인증이나 행위기반 인증에서 복제나 모방에 대응하기 위해서는 인증 문제도 anomaly detection으로 접근하는 것이 바람직하다. 그림 3은 기존의 머신러닝 기반의 (a) 2 class 인증모델과 (b) anomaly detection기반의 인증의 차이점을 보여준다.

기존의 2 class 모델에서는 복제나 모방 데이터는 others model 보다는 subject model에 가까운 것으로 분류될 수 밖에 없다.

동적전자서명인식 연구에서 숙련된 위조서명(skilled forgery)에 대응하기 위해 Auto-encoder기반 비정상 패턴탐지를 적용하였는데, 딥러닝 기반 2 class 분류기 대비 높은 성능을 얻었다[8].

한편, shared signal 기반 명의도용탐지[9]와 빅데이터 기반 비대면 본인확인도 같은 맥락에서 anomaly detection 문제로 접근할 수 있을 것이며, 피싱, 파밍 사이트의 탐지 문제에도 적용될 수 있다. 또한, 핀테크 보안 이외도 침입탐지, APT공격 탐지에도 1 class anomaly detection기술을 적용할 수 있다.

3.2. 딥러닝 기반의 회귀분석 (Regression)

회귀분석은 그동안 보안 분야에서는 많이 사용되지 않았다. 하지만 인터넷 전문은행, P2P 대출 등 핀테크 서비스가 확산되어 거래 상대방의 신뢰도를 계산하는

문제가 중요한 이슈로 떠오름에 따라 회귀분석도 중요한 보안 기술로 볼 수 있다.

대출 사기 또는 부도 위험 탐지 같은 문제는 비정상 탐지로 볼 수 있지만, 대출 이자 결정을 위한 신뢰도 분석은 비정상패턴을 탐지하는 것이 아닌 회귀분석으로 문제로 봐야 한다.

그동안 회귀분석은 주로 선형, 비선형 함수에 의존해 왔는데, 최근 딥러닝을 이용한 회귀분석의 성능이 큰폭으로 개선됨에 따라 보안분야에 적용할 회귀분석도 딥러닝을 이용해 구성할 수 있다. 보통 심층신경망은 최종단에 activation 함수로 softmax 등을 사용하는데 회귀분석에 적용할 경우, class를 1로 하고 activation 함수를 지정하지 않으면 회귀분석용으로 이용할 수 있게 된다.

향후 핀테크 서비스가 P2P 대출 등 다자간 거래로 발전할 것으로 예상됨에 따라 인터넷 전문은행 서비스에서 대출자에 대한 신용평가 뿐 아니라, 개인이 서비스의 신뢰도를 평가하거나, 타 개인에 대한 신뢰도를 평가함에 있어서 이러한 딥러닝 기반의 회귀분석 기술의 적용이 가능할 것이다.

Ⅳ. 결 론

인공지능 발전이 가속화되고 인공지능 활용이 넓어지면서, 인공지능의 보안이슈에 대해 검토하고 준비하는 것이 필요하다. 또한 딥러닝으로 대표되는 인공지능 기술을 보안 문제에 활용하는 방안도 적극적인 연구가 필요하다. 딥러닝 기반의 비정상탐지와 회귀분석은 핀테크 보안 각 분야에 적용 가능한 기술로 기존 방법에 비해 보안성 개선에 크게 기여할 것으로 보인다.

참 고 문 헌

[1] 스페셜경제, “인간의 ‘윤리’ 인공지능에게 통할까? http://blog.naver.com/speconomy/220655575503

[2] Nguyen A, Yosinski J, Clune J, “Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Image”, Computer Vision and Pattern Recognition (CVPR’15), IEEE,2015

[3] 김호광, “주목받는 딥러닝, 보안 측면의 기회와 위협”, ZDNet Korea http://www.zdnet.co.kr/column/column_view.asp?artice_id=20160316155549

[4] Wikipedia, “Anomaly Detection“, https://en.wikipedia.org/wiki/Anomaly_detection

[5] Wikipedia, “Support Vector Machine“, https://en.wikipedia.org/wiki/Support_vector_machine

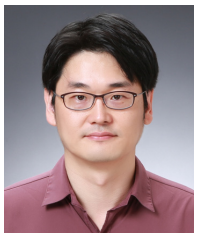
[6] S. Lange, “Deep auto-encoder neural networks in reinforcement learning”, IJCNN, 2010

[7] 채다인, 서다빈, 이영경, 최대선, “히든싱어: Auto-encoder 기반 가수 구분”, JCCI 2016

[8] 남승수, 최대선, 서창호, “숙련된 위조서명 구분이 가능한 딥러닝 기반의 모바일 동적서명 인식”, 정보처리학회 춘계학술대회 2016

[9] White paper, “The Shared Signals Model”, Open Identity Exchange

〈저자소개〉



최 대 선 (Daeseon Choi)
종신회원

1995년 2월 : 동국대학교 컴퓨터공학과 학사
1997년 2월 : 포항공과대학교 컴퓨터공학과 석사
2009년 1월 : 한국과학기술원 전산학과 박사
1997년 1월~1999년 6월 : 현대정보기술 선임
1999년 7월~2015년 8월 : 한국전자통신연구원 인증기술연구실 실장/책임연구원
2015년 9월~현재 : 공주대학교 의료정보학과 부교수
2016년 현재 : 정보보호학회 이사
<관심분야> 인증, 개인정보보호, 이상거래탐지, 의료정보보안, 머신러닝