ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

사용자 입력 패턴 및 전자 금융 거래 패턴을 이용한 모바일 뱅킹 이상치 탐지 방법☆

Outlier Detection Method for Mobile Banking with User Input Pattern and E-finance Transaction Pattern

민 희 연[1] 박 진 형[1] 이 동 훈[1] 김 인 석[1*]
Hee Yeon Min Jin Hyung Park Dong Hoon Lee In Seok Kim

요 약

모바일 뱅킹을 이용한 거래 증가세가 지속되면서 모바일 금융 보안 위협 또한 증가하고 있다. 모바일 뱅킹은 금융사가 제작한 전용 앱을 통해 금융거래를 수행하는 방식으로 인터넷 뱅킹에 준하는 대부분의 서비스를 제공하고 있다. 모바일 뱅킹 전용 앱에서 저장하고 있는 신용카드 번호와 같은 개인정보는 해커의 악의적인 공격이나 모바일 단말 분실로 인해 2차적인 공격에 이용될 수 있다. 따라서 본 논문에서는 이러한 개인정보 유출에 의한 모바일 금융사고 위협에 대응하기 위해 모바일 단말에서 뱅킹 서비스 이용 시 사용자의 입력 패턴과 거래 패턴을 이용하여 올바른 사용자에 의한 거래 시도인지 여부를 판단할 수 있는 이상치 탐지 방법을 제안한다.

사용자의 입력 패턴과 거래 패턴 데이터에는 특정 사용자를 식별할 수 있는 정보들이 포함되어 있으며, 따라서 이를 적절히 이용할 경우 올바른 사용자에 의한 금융 거래와 비정상 거래를 구분하기 위한 자료로 사용할 수 있다. 본 논문에서는 실험을 위해 스마트폰에서 직접 사용자 입력 패턴 정보를 수집하였고, 국내 모 금융사에서 이상치 탐지에 사용하는 실험 데이터를 획득하여 거래 패턴 정보로 활용하였다. 수집된 정보를 바탕으로 입력 패턴 및 거래 패턴 기반의 탐지 실험을 진행한 결과, 효율적으로 이상 거래를 탐지할 수 있음을 확인하였다.

☞ 주제어 : 이상치 탐지, 사용자 입력 패턴, 전자금융 거래 패턴

ABSTRACT

As the increase of transaction using mobile banking continues, threat to the mobile financial security is also increasing. Mobile banking service performs the financial transaction using the dedicate application which is made by financial corporation. It provides the same services as the internet banking service. Personal information such as credit card number, which is stored in the mobile banking application can be used to the additional attack caused by a malicious attack or the loss of the mobile devices. Therefore, in this paper, to cope with the mobile financial accident caused by personal information exposure, we suggest outlier detection method which can judge whether the transaction is conducted by the appropriate user or not. This detection method utilizes the user's input patterns and transaction patterns when a user uses the banking service on the mobile devices.

User's input and transaction pattern data involves the information which can be used to discern a certain user. Thus, if these data are utilized appropriately, they can be the information to distinguish abnormal transaction from the transaction done by the appropriate user. In this paper, we collect the data of user's input patterns on a smart phone for the experiment. And we use the experiment data which domestic financial corporation uses to detect outlier as the data of transaction pattern. We verify that our proposal can detect the abnormal transaction efficiently, as a result of detection experiment based on the collected input and transaction pattern data.

☞ keyword : Outlier Detection, User Input Pattern, E-finance Transaction Pattern

1. 서 론

최근 모바일 디바이스의 발전과 무선 네트워크 및 이동통신의 발전으로 모바일 디바이스 사용자 수는 지속적으로 증가해왔으며, 현재 국내 스마트폰 사용자 수는 약 3500만 명에 이르고 있다. 이에 따라 언제 어디서나 모바일 단말을 이용하여 편리하게 금융 조회 및 결제가 가능

[1] Graduate School of Information Security, Korea University, Anam-ro, Seongbuk-gu, Seoul, 136-701, Korea.
* Corresponding author (iskim11@korea.ac.kr)
[Received 24 October 2013, Reviewed 6 November 2013, Accepted 22 November 2013]
☆ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No. 2010-0020726).

한 모바일 전자금융 서비스 이용률 또한 급증하고 있다.

한국은행 보도 자료에 따르면 2013년 1/4분기 모바일 뱅킹 등록 고객 수는 전분기말(3,709만 명) 대비 10.9% 증가한 4,113만 명을 기록하였다. 이 중 스마트폰 기반 모바일 뱅킹 등록고객 수는 2,807만 명으로 전분기말(2,397만 명)대비 17.1% 늘어나면서 높은 증가세를 지속하고 있다[1].

전세계 모바일 결제 시장 규모는 2011년 이후 연간 평균 42%, 이용자 수는 평균 23% 증가하여 2016년에는 시장규모는 6,169억 달러, 이용자 수는 4억 4천만 명이 될 것으로 전망하고 있다[2]. 현재 국내 모바일 뱅킹은 대부분 iOS, 안드로이드 등 모바일 OS에 따라 금융기관이 제공하는 뱅킹 전용 앱(App)을 통해 금융거래를 수행하는 방식으로, 계좌 조회나 이체는 물론 신용카드, 펀드 등의 거래까지 인터넷 뱅킹에 준하는 대부분의 서비스를 제공하고 있다. 따라서 기존 PC환경의 인터넷 뱅킹에서 발생할 수 있는 모든 보안 위협이 모바일 뱅킹에서도 동일하게 발생할 수 있으며, 모바일 디바이스의 개방성 및 다양한 네트워크 접속환경 지원으로 인해 추가적인 보안 위협이 발생할 수 있다. 모바일 디바이스는 휴대성 및 편리성과 함께 다양한 어플리케이션의 발달로 사용자의 모든 생활에 사용되고 있어 특정 사용자의 개인 정보로 분류되는 다양한 데이터를 저장하게 된다. 특히 모바일 뱅킹 전용 앱을 이용할 경우 신용카드 번호와 같은 개인정보가 모바일에 저장되며, 이렇게 저장된 개인정보들은 해커의 악의적인 공격이나 악성코드 혹은 모바일 단말 분실로 인해 외부로 노출될 가능성이 높다[3].

「2012년 금융 스마트폰 주요 보안 이슈 및 동향 보고서」[4] 에 의하면 2010년 해외에서 금융정보를 절취하기 위한 모바일 악성코드 및 관련 해킹 툴킷이 발견되었고, 국내 연구기관 및 학회에서도 무선인터넷 접속 환경, 모바일 개방성, 개방형 모바일 애플리케이션 마켓, 도난과 분실 등 모바일 환경이 갖는 위협 요소에 따라 모바일 기반 금융보안 위협 시나리오 및 대응방안이 활발하게 발표되고 있어, 실제 모바일 금융 보안 위협의 현 상황과 이에 대한 높은 관심을 말해주고 있다. 또한 모바일 취약점을 이용한 보안 공격 위협이 점점 고도화 및 지능화, 다양화 되고 있어 개인정보 유출 및 금전적인 피해로부터 사용자를 보호하기 위한 보안기술이 요구되고 있다. 특히 민감한 정보를 다루는 금융 어플리케이션에서는 무결성, 기밀성, 가용성, 안전성, 신뢰성을 보장하기 위해 모바일 뱅킹 환경에 적합한 보안대책 관련 연구가 필요하다. 이에 본 논문에서는 모바일 단말에서 뱅킹 서비스 이용 시 사용자 입력 패턴과 거래 패턴에 기반하여 이상치(이상징후)를 탐지할 수 있는 방법을 제안한다. 사용자 입력 패턴과 거래 패턴은 개인마다 차이가 있기 때문에 이를 이용하면 정상적인 뱅킹 서비스 이용자와 공격자를 구분할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 통해 모바일 금융 보안에 대해 알아보고, 사용자를 판별하기 위한 입력패턴 기반 연구들을 살펴본다. 3장과 4장에서는 사용자 입력 패턴 및 거래 패턴에 기반한 금융거래 이상치 탐지 방법의 개념과 구조를 제안한다. 5장에서는 실험을 통해 본 논문에서 제안하는 방안의 성능과 정확성에 대해 설명하고, 6장에서 결론을 맺는다.

2. 관련 연구

2.1 모바일 전자금융 보안 대응 방법

2.1.1 모바일 뱅킹 인증

모바일 전자 금융에 대한 관심이 높아지면서 모바일 상에서의 인증에 대한 관심 또한 높아지고 있다. 사용자가 모바일 디바이스를 이용하여 온라인으로 금융거래를 수행할 경우, 기존 PC환경과는 달리 모바일 디바이스의 다양한 네트워크 접근 방법으로 인해 금융 서비스에 접근할 수 있는 경로가 늘어나게 된다. 이는 보안측면에서 볼 때 1차적으로 보호기능을 하는 인증 또한 강화되어야 한다는 것을 의미한다.

인터넷 뱅킹과 마찬가지로 모바일 뱅킹 또한 비대면 거래이기 때문에 사용자에 대한 인증과 거래승인 과정을 보호하는 것이 매우 중요하다. 이를 위해 다양한 형태의 비밀정보와 보안 프로토콜을 이용하여 사용자의 전자금융거래를 보호하고 있음에도 불구하고, 공격자는 고도화되고 지능화된 악의적인 공격방법을 이용해 거래를 자신이 의도한 형태로 변형시키는 등 현 전자 금융 보안 기술은 공격에 매우 취약하다는 것이 맹영재 등의 연구에 의해 분석된바 있다[5]. 따라서 정당한 사용자의 접근만을 허용하여 모바일 단말 분실이나 악성코드에 의한 해킹으로부터 모바일 금융거래를 안전하게 보호하는 인증 기술에 대한 연구가 지속적으로 필요하다.

최근에는 모바일 디바이스의 간편한 휴대성으로 인해 항시 소지가 가능하다는 점을 이용하여 모바일 디바이스를 인증매체로 이용하거나 혹은 모바일 기기에 내장된 센서를 이용하여 인증기술을 구현하는 방식의 연구가 주를 이루고 있다. 또한 생성된 인증정보를 서버로 안전하

게 전송하기 위한 프로토콜에 대한 연구도 활발하게 진행되고 있으며, 생체정보 인식 기술 뿐만 아니라 생체정보를 이용한 사용자 인증값 추출과 안전한 인증값 저장에 대한 연구도 진행되고 있다[6].

2.1.2 이상 금융거래 탐지

이상금융거래탐지는 전자금융거래환경에서 사용자의 이용환경, 거래패턴, 거래사전행위에 의해 종합적으로 비정상적인 금융거래를 판별하는 것이다. 이용환경은 하드웨어, 소프트웨어, 네트워크 등을 통해 고유정보변화를 탐지하고 거래패턴은 금융거래금액, 거래횟수, 거래계좌 등 거래관련 정보를 기반으로 하며, 비정상 사전행위는 전자 금융 거래를 위해 필요한 사용자의 사전 행위를 통해 판별하게 된다. 사용자의 과거 금융거래 정보를 수집하여 분석하면 일정한 범위와 형태로 패턴이 만들어질 수 있는데, 이 패턴을 토대로 새롭게 이용되는 금융거래에서 금융거래행위가 미리 수집된 사용자의 사전행위의 정상적인 범위 계도를 벗어나게 되면 탐지 시스템을 통하여 이상금융거래로 판별할 수 있다. 이상금융거래 탐지에서는 과거 정상 행위들을 통해 비정상 거래 발생 시 이를 탐지하는데 있어 정확률(precision)을 높이는 것이 중요하며 이상금융거래 탐지 여부에 따라 사용자 금융거래 정보는 지속적으로 재설정 되어야 한다[7].

2.2 입력 패턴 기반 사용자 인증 연구

모바일 디바이스에서 사용자 입력 패턴에 기반한 사용자 인증과 관련된 연구는 국내외에서 모두 진행 중에 있으며 국내보다는 국외에서 더 활발히 진행되고 있다. 사용자 입력 패턴을 이용한 사용자 인증 관련 연구는 모바일 디바이스의 특성으로 인해 화면 터치로 입력되는 터치 시간이나 압력 등의 패턴이 개인별로 뚜렷하게 차이가 있으므로 이를 모니터링 한 후 정상적인 사용자와 공격자를 구분하는 방향으로 진행되고 있다.

Alexander De Luca 의 연구[8]에서는 터치 스크린 패턴에서 발생할 수 있는 숄더서핑 및 스머지 공격, 무작위 대입 공격에 대한 방안으로 스크린 패턴 입력 시 발생하는 시간, 압력, 속도, 좌표 데이터를 Dynamic time warping(DTW) 알고리즘을 이용하여 사용자를 인증하는 방법을 제안하였다. 그러나 터치 스크린 패턴 자체의 데이터 양이 워낙 작기 때문에 실험을 위한 현실적인 데이터의 양을 수집하는데 한계가 있어 정확한 실험결과를 기대하기 어렵고, 패턴이 바뀔 때마다 사용자의 데이터가 바뀌기 때문에 저장해야하는 데이터가 많아진다는 단점이 존재한다.

서호진 등은 사용자 입력 패턴 중 터치와 관련된 압력, 면적 및 스크롤과 관련된 시간, 압력, 면적 등의 입력 값을 역전파 신경망(Back Propagation Neural Network)의 입력 노드로 사용하여 이상 징후를 탐지하였다[9]. 그러나 사용자 입력 패턴 데이터 수집 시 멀티터치를 고려하지 않고 싱글터치만 고려하였으며, 이는 모바일 뱅킹 거래 시 두 개의 손가락을 동시에 사용하게 될 경우 입력패턴 정보가 달라져 정상적인 거래임에도 인증 단계에서 의심스러운 거래로 탐지될 수 있다. 또한 신경망 알고리즘을 사용하는 경우 정확도를 높이기 위해 은닉노드의 수를 많이 늘려야 하나 은닉노드가 많을수록 부하가 많이 걸리며 은닉노드 부분이 공개되지 않아 신뢰성이 떨어진다는 단점이 있다.

Yuxin Meng의 연구[10]에서는 싱글 터치, 멀티 터치, 스크롤의 시간, 각도, 스피드, 압력 등의 입력 값에 대해 PSO-RBPN 알고리즘을 사용하여 사용자 인증을 하였다. 이 연구에서는 사용자의 입력 값을 싱글 터치 뿐 아니라 멀티 터치까지 확장하여 함께 고려했다는 장점이 있다. 멀티 터치 시에는 더 많은 이벤트가 동시에 발생하기 때문에 해당 이벤트 값을 이용하면 싱글 터치에 비해 사용자를 좀 더 정확히 구분해낼 수 있다. 하지만 입력 값 측정을 위한 학습시간을 10분으로 잡고 실험하였기 때문에 실제 모바일 금융거래를 이용하는 고객에게 위 방법을 적용할 경우 편리성과 효율성이 많이 떨어질 수 있다.

기존의 입력 패턴을 이용한 인증 연구들은 오탐률을 줄이고 빠른 시간에 학습을 할 수 있는 효율적인 방안을 제시하였지만 사용자 입력 패턴은 항상 동일할 수 없기 때문에 사용자 입력 패턴에만 의존한 탐지 방법은 낮은 확률이라 하더라도 인증과정에서 오탐이 발생할 수 있다. 따라서 본 논문에서는 기존 사용자 입력 패턴 기반 탐지 방법을 효율적으로 개선하고, 추가적으로 오탐 확률을 줄이기 위해 거래 패턴을 이용하는 이상 징후 탐지 방법을 제안한다.

2.3 Support Vector Machine(SVM)

특정 사용자의 패턴을 구별하기 위한 데이터마이닝 방법으로는 클러스터링 기법과 같은 비교사 학습 방법 보다 목표 변수가 설정된 Classification 기법에 해당하는 교사 학습 방법이 더 적합하다. 왜냐하면 금융 서버로 전송

되는 사용자 패턴 정보에는 어떤 사용자의 패턴 정보인지 목표 변수를 설정할 수 있기 때문이다. 이러한 Classification 기법 중 SVM(Support Vector Machine)은 학습 데이터의 수가 적은 경우에도 Support Vector 를 이용한 분류 경계면(Hyperplane)을 형성하기 때문에 높은 분류 정확도를 보여준다. SVM은 일반적으로 이진분류 학습방법으로 벡터 공간 내에 표현된 입력 패턴 중 평면에 가장 가까운 두 개의 서포트 벡터의 거리를 최대화하는 최대 마진 분류 평면을 찾음으로써 데이터를 두 개의 클래스로 분류한다. 가장 초기의 최적화된 Hyperplane은 선형분류기였으나 이후 Bernhard Boser 등에 의해 Kernel Trick을 응용하는 방법이 제안된 후, 비선형 Kernel Function을 이용하여 입력값을 매핑되는 Feature Space에 적합하게 구성할 수 있게 되었다. SVM의 경우 아무리 많은 학습 데이터가 제공된다고 하더라도 결국 최종 모형은 소수의 서포트 벡터만을 이용해서 도출되기 때문에, 학습에 많은 데이터를 필요로 하지 않아 데이터 수집에 시간을 줄일 수 있다는 특징을 갖는다. 또한 분류 인식 성능과 처리속도가 매우 뛰어나기 때문에 최근에 패턴분류 및 예측방법으로 각광받고 있다. 또한 가지분할을 위한 변수 선택에 따라 결과가 달라진다는 의사결정나무의 한계점과 인공신경망의 과적합, 국소최적화와 같은 한계점들을 완화한다는 장점을 갖고 있다[11,12,13].

본 논문의 이상 징후 탐지 프로세스는 가용성과 편리성, 즉시성이 중요한 금융 거래를 위한 모델이다. 따라서 모델을 생성하는 시간이 빠르며, 적은 학습 데이터에도 불구하고 사용자의 입력 패턴을 높은 정확도를 가지고 구분할 수 있는 SVM 알고리즘을 이용하여 사용자의 이상 입력 패턴을 탐지한다.

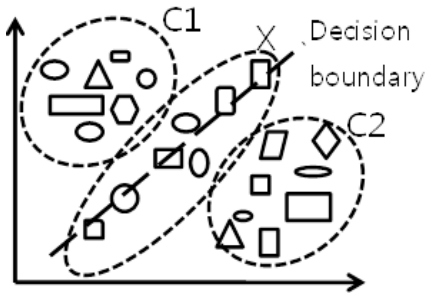
2.4 상자 그림(Box Plot) 모델

Box Plot 은 두 개 이상의 집단 혹은 변수에서 최대값(max)과 최소값(min) 그리고 중앙값(자료를 크기순으로 나열했을 때 가운데 위치하는 값: median) 및 사분위수(자료를 크기 순서에 따라 늘어놓은 자료를 4등분 했을 때 위치하는 값을 의미함) 중 제 1사분위수(아래에서 25% 백분위점에 위치하는 수: Q1), 제 3사분위수(아래에서 75% 백분위점에 위치하는 수: Q3)의 다섯 숫자를 요약하여 그래프로 나타내는 방법으로 John W. Tukey가 제안한 탐색적 데이터 분석 방법이다. 이 방법은 데이터를 정규분포로 나타내어 분석하고자 하는 데이터 패턴을 한눈에 볼 수 있으며, 데이터의 위치에 따라 이상치를 정확히 알 수 있다는 장점이 있다[14,15,16].

3. 제안하는 방법

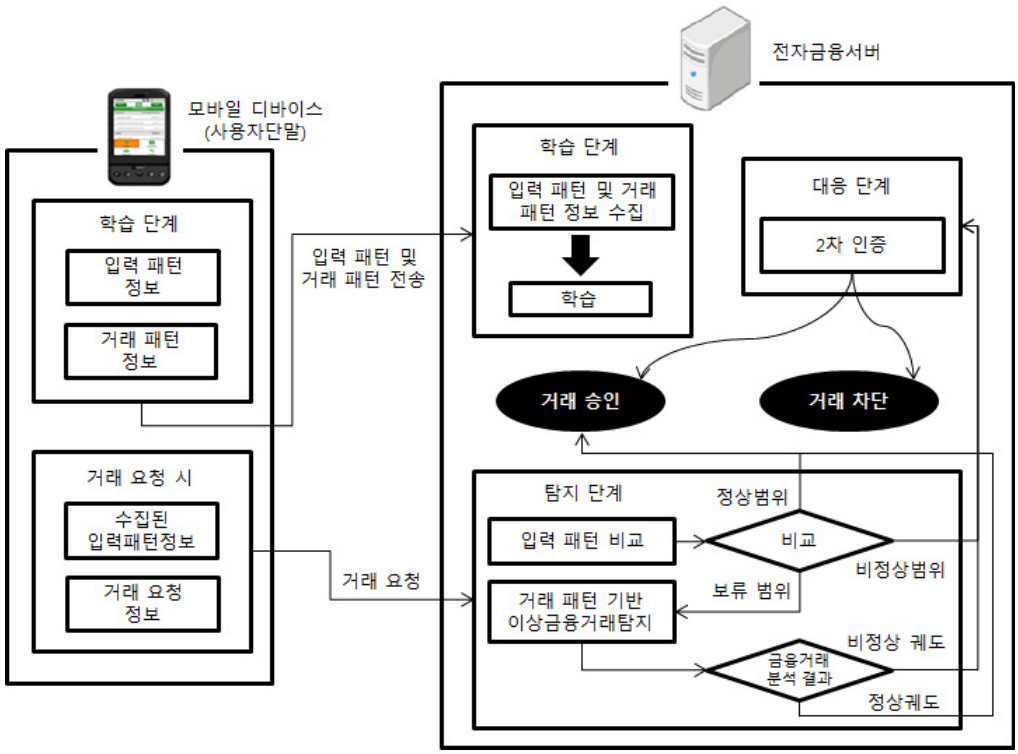
본 장에서는 모바일 디바이스의 입력 패턴 및 전자금융 거래 패턴을 이용하여 이상 거래를 탐지하는 방법에 대해 설명한다. 모바일 디바이스의 입력 패턴은 사전(거래 요청 이전) 탐지에 사용되며, 전자 금융 거래 패턴은 사후(거래 요청 이후) 탐지에 사용된다. 사용자의 입력 패턴과 거래 패턴은 모두 사용자 개개인의 특성(패턴)이 반영되기 때문에 이를 연동할 경우, 더욱 정확한 탐지가 가능하다.

3.1 개념



(그림 1) 불확실성을 내재하고 있는 데이터 분류 기법
(Figure 1) The uncertainty inherent in a data classification technique

터치스크린 기반의 모바일 디바이스를 이용한 금융 거래의 경우 터치스크린으로 입력되는 사용자의 입력 패턴 및 금융 거래 데이터 정보에 따라 사용자 별로 일정한 패턴이 형성된다. 이러한 패턴 정보를 이용하면 (그림 1)과 같이 정상 사용자 패턴과 비정상 사용자 패턴으로 구분할 수 있다. (그림 1)에서 C1 그룹을 정상 사용자의 패턴이라 보고 C2 그룹을 비정상 사용자의 패턴이라고 볼 때, 이들을 구분할 수 있는 분류경계를 구할 수 있다면 이를 바탕으로 데이터들을 정상/비정상으로 분류하여 탐지를 할 수 있다. 이 때, 그림 1에서 X범위에 있는 데이터와 같이 경계 부근에 속하는 데이터들의 경우 정확한 분류가 어려울 수 있으며, False Negative 혹은 False Positive 같은 오분류를 발생 시킬 수 있다. 금융거래는 사용자 측면에서 가용성이 매우 중요하기 때문에 만약 금융기관이 사용자 입력패턴 인증을 실제 모바일 뱅킹 환경에 적용할



(그림 2) 입력 패턴과 거래 패턴을 이용한 이상 징후 탐지 모델

(Figure 2) The anomaly detection model using both user input pattern and e-finance transaction pattern

경우, 단 1명이라도 인증 오류에 의한 거래 차단이 발생하는 상황 혹은 비정상 사용자를 정상 사용자로 판단함으로 인해 발생하는 금융 관련 사고는 금융기관에 큰 부담이 될 수 있다. 따라서 모바일 금융 거래 시 X의 범위의 데이터가 발생시킬 수 있는 오분류율을 최대한 줄이는 것이 매우 중요하다.

본 논문에서는 스마트폰과 같은 모바일 디바이스에서 터치스크린을 통한 사용자의 입력 패턴 정보와 거래 패턴 정보를 이용하여 정상 사용자와 비정상 사용자의 전자 금융 거래 시도를 구분하여 비정상 사용자의 거래 시도를 탐지한다. 이러한 탐지는 높은 정확성과 실시간성이 보장되어야 한다. 입력 패턴을 이용한 탐지 과정에서 적은 입력패턴정보, 입력패턴 정보에 포함된 잡음, 사용자의 모바일 뱅킹 부적응 등으로 인하여 정확한 사용자 입력패턴을 탐지하는데 오분류가 발생할 수 있다. 이러한 오분류를 일으킬 수 있는 특정 데이터에 대해서, 사용자의 전자 금융 거래 패턴 기반의 Box Plot 모델을 이용한 이상 금융 거래 탐지 과정을 한 단계 더 거치면 오분류를 최소화할 수 있다. 통계적으로 사용자의 전자 금융 거래는 일정한 패턴을 보이게 되므로 이러한 패턴 정보를 이상 금융 거래 탐지에 사용될 수 있다.

3.2 탐지 모델

제안하는 이상 징후 탐지 모델은 (그림 2)와 같이 크게 사용자의 모바일 디바이스와 전자금융서버의 두 개체를 통해 이루어지며 학습과 탐지, 대응의 세 단계로 구성된다.

학습 단계에서는 전자금융서버가 모바일 디바이스로부터 전송된 사용자의 입력패턴을 수집한 후 SVM 알고리즘을 이용하여 사용자 입력 패턴 데이터에 대한 학습을 진행하고, 금융 서버에서 수집된 특정 기간 동안의 사용자 거래 패턴 데이터를 이용해 Box Plot 모델을 만든다. 탐지 단계에서는 거래 요청 시 발생하는 사용자 입력 패턴 데이터와 금융 거래 요청 정보를 이용하여 기존 학습된 데이터와 비교함으로써 모바일 뱅킹 시 발생한 거래 요청이 정상 사용자에 의한 거래 요청인지 확인한다. 탐

지 단계는 2단계로 구성되는데, 첫 번째 단계에서는 사용자의 입력 패턴 정보를 기반으로 분석을 수행한다. 기존 학습된 입력 패턴 정보에 따라 거래 요청 시 발생한 패턴 데이터가 정상으로 분류될 경우 금융 거래를 승인하고, 비정상으로 분류될 경우 거래를 승인하지 않고 대응 단계로 넘어간다. 보류로 분류될 경우에는 탐지의 두 번째 단계로 진행되며, 요청된 거래 패턴을 이용하여 추가적인 분석을 수행한다. 이 때, 학습 단계에서 수립된 Box Plot 모델에 따라 비정상 궤도로 분류될 경우 대응 단계로 넘어간다.

대응단계에서는 탐지 단계에서 비정상 거래 요청으로 판별된 경우에 대해 추가적인 인증 방식을 사용하여 2차 인증 절차를 진행한다. 2차 인증을 통과하는 경우 거래 요청을 승인하고, 인증을 통과하지 못할 경우에는 요청된 금융거래를 차단한다.

4. 제안하는 기법 설계

4.1 정보 수집 단계

사용자가 터치스크린 기반의 모바일 디바이스를 이용할 때 터치스크린으로부터 받을 수 있는 입력 패턴은 크게 터치와 스크롤로 구분할 수 있으며, 스크롤은 다시 손가락 하나를 이용한 단일 스크롤과 여러 개의 손가락을 이용한 멀티 스크롤로 구분할 수 있다. 터치는 버튼을 누르거나 비밀번호 입력을 위해 가상 키보드를 손가락으로 눌렀다 떼는 입력이고, 스크롤은 하나 또는 두 개의 손가락으로 스크린을 누른 채로 움직이는 입력이다. 터치 및 스크롤과 같은 사용자 입력 패턴은 신체적 특징이나 행동 성향에 따라 사용자마다 다르게 나타나기 때문에 이를 적절히 이용하면 사용자의 정상 패턴과 대비되는 이상 패턴을 탐지할 수 있다. (표 1)은 모바일 디바이스로부터 획득할 수 있는 사용자 입력 패턴 값의 종류를 나타낸 것이다.

터치는 사용자가 패스워드 입력을 위해 화면에 나타나는 키보드를 누르는 경우의 입력 패턴을 수집하여 활용한다. 터치 입력 시에는 누르는 시간과 압력, 그리고 누르는 크기를 측정할 수 있다. 패스워드와 같은 사용자의 비밀정보 등을 입력하는 경우에는 개인의 친숙도 및 숙련도에 따라 수집된 입력 패턴 결과에 큰 차이가 있을 수 있다. 따라서 한 번의 터치 입력과 다음 터치 입력까지의 시간이나 전체 입력 시간 등은 사용자를 구분할 수 있는 주요 정보로 사용할 수 있다.

(표 1) 모바일 디바이스에서 수집 가능한 사용자 입력 값
(Table 1) Collectible user input pattern data in the mobile device

<table>
<tr><td colspan="2">분류</td><td colspan="2">사용자 입력 정보</td></tr>
<tr><td colspan="2" rowspan="5">Touch</td><td colspan="2">터치 압력</td></tr>
<tr><td colspan="2">터치 크기</td></tr>
<tr><td colspan="2">터치 시간</td></tr>
<tr><td colspan="2">터치 입력 사이의 시간 간격</td></tr>
<tr><td colspan="2">전체 입력 시간</td></tr>
<tr><td rowspan="11">Scroll</td><td rowspan="6">Single</td><td>스크롤 시작 좌표</td><td rowspan="2">스크롤 기울기</td></tr>
<tr><td>스크롤 끝 좌표</td></tr>
<tr><td>스크롤 시간</td><td rowspan="2">스크롤 속도</td></tr>
<tr><td>스크롤 길이</td></tr>
<tr><td colspan="2">스크롤 시작 압력 및 크기</td></tr>
<tr><td colspan="2">스크롤 끝 압력 및 크기</td></tr>
<tr><td rowspan="5">Multi</td><td colspan="2">첫 번째 터치 손가락 및
두 번째 터치 손가락 구분</td></tr>
<tr><td>첫 번째/두 번째 터치
스크롤 시작 좌표</td><td rowspan="2">첫 번째/두 번째
터치 손가락
스크롤 기울기</td></tr>
<tr><td>첫 번째/두 번째 터치
스크롤 끝 좌표</td></tr>
<tr><td>첫 번째/두 번째 터치
스크롤 시간</td><td rowspan="2">첫 번째/두 번째
터치 손가락
스크롤 속도</td></tr>
<tr><td>첫 번째/두 번째 터치
스크롤 길이</td></tr>
<tr><td colspan="2">첫 번째/두번째 터치 시작 압력, 크기</td></tr>
<tr><td colspan="2">첫 번째/두번째 터치 끝 압력, 크기</td></tr>
</table>

스크롤은 사용자가 화면을 이동하거나 확대·축소하기 위해 하나 또는 두 개의 손가락을 이용해 스크린을 누른 상태로 움직이는 경우의 입력 패턴을 수집한다. 수집되는 정보는 터치의 경우와 마찬가지로 시간, 압력, 크기를 비롯하여 하나의 손가락을 이용하는 싱글 터치의 경우, 스크롤의 시작 및 끝 좌표 값과 스크롤 길이가 포함된다. 멀티 터치의 경우에는 터치되는 손가락의 순서로 각 손가락을 구분할 수 있으며, 각 손가락에 대해 스크롤 시작 및 끝 좌표 값과 스크롤 길이가 포함된다. 멀티 스크롤 입력 시 터치스크린에 손가락이 닿는 순서대로 첫 번째 입력 손가락과 두 번째 입력 손가락이 구분되는데, 이러한 정보를 수집된 스크롤 패턴 정보와 결합하면 사용자를 구분할 수 있는 주요 요소가 된다.

입력 패턴과 마찬가지로 사용자의 금융 거래 패턴 또한 사용자 별로 구분되는 일정한 패턴을 보인다. 거래 시간이나 거래 금액, 거래 위치 등의 금융 거래 정보를 이용하면 기존의 정상 금융 거래 패턴과 대비되는 이상 거래 패턴을 탐지할 수 있다. 사용자의 금융 거래 과정에서 전자 금융 서버는 로그인 정보(시간, 위치), 거래 전 사전행

위, 거래일자, 거래시간, 거래일련번호, 서비스구분, 거래종류, 계좌번호, 실명번호, 채널유형, 거래단말번호, 거래점번, 거래금액, 거래 이용 빈도수, 이체 대상 계좌번호 등 다양한 거래 정보를 기록하고 수집할 수 있다.

4.2 학습 단계

학습 단계에서는 금융서버가 정보 수집 단계에서 수집된 사용자 입력 패턴 정보와 거래 패턴 정보를 이용하여 탐지 모델을 구성한다.

사용자 입력 패턴을 기반으로 모델을 구성할 때에는 먼저 학습 알고리즘에 입력되는 입력 패턴 정보들에 대해 Feature Selection 과정을 거쳐 정확도는 그대로 유지하면서 모델 생성 시간을 줄일 수 있도록 사용자 구분에 변별력을 갖는 속성을 추출한다. 속성들이 추출되면 해당 속성들을 이용하여 학습을 실시한다. 본 논문에서는 학습 알고리즘으로 Classification 방법들 중 SVM(Support Vector Machine) 알고리즘을 이용하였다. SVM 알고리즘은 Support Vector를 이용하여 분류 경계면을 형성하기 때문에 많은 학습 데이터가 제공된다고 하더라도, 결국 최종 모형은 소수의 Support Vector만을 이용해서 도출된다. 따라서 학습에 많은 데이터를 필요로 하지 않아 정확한 분류를 위해 필요한 데이터 수집 시간을 줄일 수 있다는 장점이 있다. 또한 분류 인식 성능과 처리 속도가 매우 뛰어나 금융 거래에서 이상 거래 패턴 탐지를 위한 패턴 정보 학습에 사용하기 적합하다. 일반적인 SVM은 기본적으로 이진 분류기로 동작하며, 본 논문에서 사용하는 WEKA 라이브러리의 SMO(Sequential Minimal Optimization) 또한 기본적으로 이진 분류기로 동작한다. 따라서 이러한 이진 분류기를 이용하여 Multi-Class 분류를 하기 위해서 One-versus-All, One-versus-One 접근 방식을 이용할 수 있다[18]. One-versus-All 방식은 입력 데이터들을 하나의 Label을 갖는 그룹과 나머지 모든 Label을 갖는 그룹으로 나누어 순차적으로 학습을 진행한다. 이러한 그룹 분류는 Label을 갖는 모든 그룹이 다른 모든 그룹과 비교될 수 있도록 반복하여 진행되며, 이를 이용하여 최종 모델을 생성한다. 이후 분류 시에는 Winner-takes-all 전략을 이용한다. One-versus-One 방식은 본 논문에서 사용된 방식으로써, 입력 데이터의 모든 Label 쌍에 대하여 학습을 진행하여 모델을 생성하며 분류 시 Max-wins voting 전략을 이용한다.

거래 패턴 기반 탐지를 위한 Box Plot 모델로는 Typical Schematic Box Plot을 사용하는데, 이는 직관적으로 어떤 데이터가 이상치(outlier)일 가능성이 큰지 쉽게 알 수 있도록 표현한다. Box Plot 모델 생성 과정에서는 사용자의 거래 시간, 거래 위치(IP정보), 거래 사전 행위 정보를 이용하여 학습을 진행한다. 학습 과정은 다음과 같다. 가장 먼저 정보 수집 단계에서 수집된 위 3가지 정보(거래 시간, 거래 위치, 거래 사전 행위정보)를 이용하여 백분위 값을 구한다. 백분위 값을 토대로 각각의 패턴 데이터 백분위 통계를 구하면 각각의 패턴에 대한 백분위 평균으로 나올 수 있는 모든 가지 수를 구할 수 있다. 이 데이터들은 사용자에 대한 정상 패턴의 자료들이 된다. 따라서 비정상 거래 데이터를 추출하기 위해 전체 즉, 1의 값에서 정상 패턴의 값을 뺀 백분위의 값을 오름차순으로 정렬하여 데이터 리스트를 구성한다.

구성된 전체 데이터 리스트 n개의 데이터에 대해 아래의 식으로 최대값(Max), 최소값(Min), 1사분위수(Q1), 3사분위수(Q3), 중간값(Median), IQR(Inter-Quartile Range) 을 구한다.

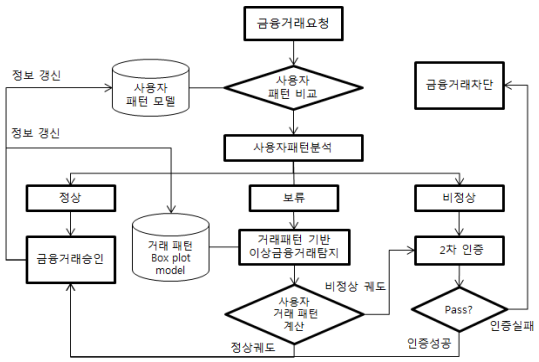
- n : 전체 데이터 수,
  데이터 리스트 X : {X0, X1, ... , Xn} 일 때,
- 최소값(Min) = X0
- 최대값(Max) = Xn
- 중간값(Median) = (n/2)th value
- 1사분위수(Q1) = $\frac{1}{4}(n+1)^{th}$ value
- 3사분위수(Q1) = $\frac{3}{4}(n+1)^{th}$ value
- IQR(Inter-Quartile Range) = Q3 - Q1

구해진 6가지 값을 이용하여 Box Plot 모델을 구성한다.

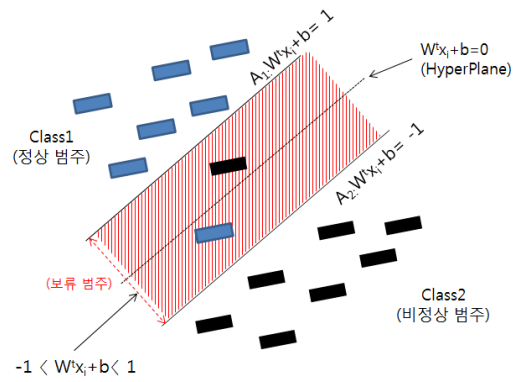
사용자의 입력 패턴 및 거래 패턴은 시간이 지남에 따라 변할 수 있기 때문에(서비스 이용자의 패턴 변화), 이후 탐지 과정을 통해 정상 패턴으로 분류되어 거래가 승인된 경우에는 해당 데이터를 이용하여 지속적으로 재학습해야한다.

4.3 탐지 및 대응 단계

사용자별 입력 패턴 및 거래 패턴에 대한 학습이 완료되면 이후 사용자가 모바일 뱅킹 서비스를 이용할 때, 서버로 전송된 사용자의 입력 패턴과 거래 요청 데이터를 학습된 모델에 입력하여 결과가 동일한 사용자의 패턴으로 분류되는지 확인함으로써 이상 거래 여부를 탐지할



(그림 3) 전자 금융 서버의 이상 징후 탐지 프로세스
(Figure 3) The abnormal detection process of the e-financial server



(그림 4) Linear SVM의 패턴 분류 범주
(Figure 4) The pattern classified category of Linear SVM

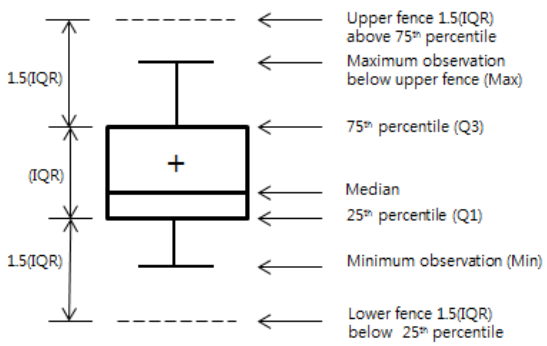
수 있다.

(그림 3)은 전자 금융 서버에서의 이상 징후 탐지 과정을 나타낸다. 사용자가 금융 거래를 요청하면 서버는 모바일 디바이스로부터 사용자 입력패턴 정보와 거래 요청 데이터를 전송받게 되는데 이때 사용자 입력패턴 정보를 이용한 탐지를 먼저 진행한다. 거래 요청 시 모바일 디바이스에서 사용자 입력패턴 정보를 수집한 후 이를 서버에 전송하면, 서버는 전송받은 사용자 입력 패턴 정보를 학습된 SVM 모델에 입력하여 분류 결과 값이 해당 사용자로 분류되는지 확인한다. 이 때, 사용자 입력패턴 정보를 기존 학습된 패턴 정보와 비교하여 3가지 범주 - 정상 패턴(정상 범주), 비정상 패턴(비정상 범주), 분류경계 범위 패턴(보류 범주) - 로 구분한다. (그림 4)는 분류 범주를 Linear SVM의 예를 이용하여 간략하게 도식화 한 것이다. (그림 4)에서 W는 법선 벡터, x_i는 학습을 위한 입력 데이터를 나타내며, 파란색 네모는 Class 1으로 구분된 데이터, 검은색 네모는 Class 2로 구분된 데이터를 나타낸다. 일반적으로 SVM 알고리즘에 의해 모델링이 될 때에는 Support Vector가 정해지고 Support Vector 사이의 최대 마진을 갖는 초평면(Hyperplane)이 결정된다. 이 과정에서 법선벡터 W와 b값이 정해지게 되는데, 추후 입력 값을 분류할 때, $W^tx_i + b$를 이용하여 그 값이 0보다 큰 경우에는 Class 1으로 분류시키고, 0보다 작은 경우에는 Class 2로 구분시킬 수 있다. 즉, 해당 초평면을 기준으로 입력값을 분류하는 것이다. 그러나 이 때, False Positive의 문제가 발생할 수 있다. SVM에 의해 설정된 $-1 < W^tx_i + b < 1$ 사이의 공간은 기존 학습에 사용된 입력 데이터들이 존재하지 않는 공간이다. 따라서 $W^tx_i + b = 0$으로 존재하는 초평면에 의해 데이터가 오분류 될 수 있다. 예를 들어, 어떤 입력 데이터가 벡터 공간에서 $0 < W^tx_i + b < 1$에 속하여 Class 1으로 분류되었지만 실제로는 Class2 의 데이터 일 수 있다. 즉, $-1 < W^tx_i + b < 1$ 공간에 속하는 데이터들은 기존 분류에 사용되었던 학습 데이터들의 Support Vector 외부에 위치하기 때문에 초평면을 기준으로 분류한다고 하더라도, False Positive의 가능성을 내재하고 있게 된다. 따라서 학습 데이터에 의해 분류된 $-1 < W^tx_i + b < 1$ 공간을 분류경계 범위(보류 범주)로 설정하고, 해당 입력 데이터 요청 시점의 거래 패턴 데이터들에 대해 Box Plot을 이용한 추가적인 분석 과정을 수행함으로써 전체적인 오분류율을 줄일 수 있다.

SVM에 의한 비교 결과가 정상 패턴에 포함되면 금융거래를 승인하고, 비정상 패턴으로 분류되면 대응 단계로 넘어간다. 결과가 분류경계 범위 패턴에 포함된 경우에는 거래 패턴을 이용하여 추가적인 탐지 과정을 진행한다.

즉, 사용자의 입력 패턴이 확실하게 정상, 비정상 범주에 속하는 경우에는 금융거래를 계속 진행하거나 혹은 이상 징후에 대한 대응을 진행하며, 패턴 비교 값이 모호한 경우(분류경계범위)에는 거래 패턴을 이용한 추가적인 탐지 과정을 진행함으로써 전체적인 탐지 과정의 효율성과 정확도를 높일 수 있다.

거래 패턴을 이용한 탐지는 사용자 입력 패턴 탐지 과정과 마찬가지로 거래 관련 데이터를 기 학습된 Box Plot 모델에 적용하여 이상치를 탐지한다. (그림 5)는 학습 단계에서 생성된 Box Plot 모델과 탐지 범위를 나타낸다. 학습 단계에서 구한 6가지 - 최대값(Max), 최소값(Min), 1사분위수(Q1), 3사분위수(Q3), 중간값(Median), IQR(사분위



(그림 5) Box Plot 모델
(Figure 5) The Box Plot model

수 범위) - 값에 의해 구성된 모델에 따라 이상치를 판단할 수 있다. 일반적으로 Box Plot 모델은 Upper fence 이상의 값과 Lower fence 이하의 값으로 분류되는 데이터를 이상치로 판단하는데, (그림 5)의 모델에서는 Q3 + (1.5 * IQR) 이상인 Upper fence 위의 값과 Q1 - (1.5 * IQR) 이하인 Lower fence 아래의 값에 해당하는 데이터를 이상치로 판단하게 된다. 이상치를 판단하는 척도는 모델을 적용하는 각 분야의 데이터 및 정책을 분석하여 그에 맞게 2.0 * IQR 혹은 3.0 * IQR 등 IQR값에 곱해지는 변수 값을 지정함으로써 조정이 가능하다. 즉, 정상 데이터와 비정상 데이터 구분을 위한 임계치는 모델이 적용되는 데이터 및 환경에 따라 설정될 수 있다. Box Plot 모델에 따른 분석 결과가 정상 범위 이내일 경우에는 금융 거래를 승인하고, 범위를 벗어나 비정상 데이터로 판단되는 경우에는 대응 단계로 넘어간다.

대응 단계는 사용자 입력 패턴을 이용한 탐지 과정에서 비정상 입력 패턴으로 분류되거나, 거래 패턴을 이용한 Box Plot 모델 분석 과정에서 비정상 궤도로 분류되어 이상 거래로 판단되는 경우 수행된다. 대응 단계에서는 SMS나 ARS와 같은 방식의 추가적인 인증 방식을 사용하여 2차 인증 과정을 수행한다. 2차 인증을 통과하는 경우 거래 요청을 승인하고, 인증을 통과하지 못했을 경우 해당 거래 요청을 차단한다.

5. 실험 및 평가

본 장에서는 사용자 입력 패턴을 이용한 이상 금융거래 탐지에 대한 실험을 진행하여 탐지 방법의 효율성과 정확도를 측정하였다. 또한 특정 사용자의 거래 시간, 거래 위치(IP정보), 거래 사전 행위 정보를 이용하여 Box Plot 모델을 구성해보았다.

(그림 6) 입력 패턴 수집을 위한 앱 구성 화면
(Figure 6) The basic screen of application for input pattern collecting

5.1 실험 환경

본 실험에서는 안드로이드 운영체제(4.1.2 버전) 기반의 스마트폰(Galaxy Nexus)을 이용하여 사용자 입력 패턴을 수집하였다. 입력패턴 정보를 수집하기 위한 앱을 제작하여 스마트폰에 설치하였으며, 해당 앱은 Android API 중 MotionEvent 클래스와 GestureDetector 클래스를 사용하여 입력 패턴 정보를 수집한다.

터치 입력 패턴을 수집하기 위해 버튼 위젯을 사용해 가상 키보드를 구성하였고, 터치 입력 후 싱글터치 스크롤 입력 패턴 수집과 멀티터치 스크롤 입력패턴 수집을 순차적으로 수행하도록 구성하였다. 앱은 (그림 6)과 같이 구성하였으며, 앱에서 수집하는 입력 패턴 정보는 4.1절의 (표 1)과 같다. 터치 입력에서는 각 터치 입력의 압력, 크기, 터치 시간, 각 터치 입력간의 시간 간격, 전체 입력 시간의 5가지 정보를 수집하였으며, 싱글 스크롤 입력에서는 스크롤 시간, 스크롤 길이, 스크롤 시작시 정보(좌표, 압력, 크기), 스크롤 종료시 정보(좌표, 압력, 크기)의 8가지 정보를 수집하였다. 멀티 스크롤 입력에서는 두

개의 손가락이 이용되기 때문에, 각 손가락에 대해 싱글 스크롤 입력에서 수집되는 정보를 수집하여 총 16가지 정보를 수집하였다. 따라서 수집된 총 입력 패턴 정보는 29가지이다. 이 중 스크롤 좌표 정보는 입력 시마다 그 값이 달라질 수 있다. 따라서 시작, 종료 좌표 값은 직접 사용하지 않고 기울기를 계산하여 기울기 값을 입력 패턴 정보로 사용한다. 즉, 학습을 위한 사용자 입력 패턴 정보는 총 26가지가 사용된다.

실험 데이터 수집은 총 60명을 대상으로 진행하였으며 20~50대까지 각 연령대 별로 스마트폰 사용 경험이 있는 15명에 대해 데이터를 수집하였다. 데이터 수집횟수는 인당 20회로 총 1200개의 데이터를 수집하여 학습 및 탐지에 활용하였으며, 각 데이터의 Class Label은 수집 대상 사용자 이름의 영문 이니셜을 사용하였다.

수집된 사용자 입력 패턴을 학습하기 위한 알고리즘으로는 SVM을 사용하였으며, 실험을 위한 툴로는 무료 데이터 마이닝 툴인 WEKA[17]를 사용하였다. WEKA는 데이터 전처리, 분류, 회귀, 클러스터링, 연관 규칙, 시각화 등 다양한 데이터마이닝 툴을 포함하고 있는 오픈소스 프레임워크로 GUI 환경을 제공한다. WEKA에서 제공하는 SVM 알고리즘은 SMO(Sequential Minimal Optimization)로, SVM 학습을 진행하는 동안 최대의 여백(Margin)을 구하는 최적화 문제를 해결하기 위한 알고리즘이다. 1998년 John Platt에 의해 제안되었으며 현재 SVM 학습을 위해 널리 사용된다. SMO 알고리즘의 주요 변수로는 C(Complexity parameter), 커널 함수 종류, 커널 함수 파라미터, Epsilon(for round-off error)의 4가지 종류가 있다. 실험은 SMO 알고리즘의 입력 변수인 커널 함수 종류 및 커널 함수 파라미터를 변경해가며 실험을 진행하였으며, 3.0 GHz CPU와 4GB 메모리를 가진 PC 환경에서 진행하였다. 추가로 다른 알고리즘과의 성능 및 정확도의 비교를 위해 타 연구에서 진행했던 실험 환경과 동일한 컴퓨팅 환경을 구성하여 실험을 진행하였다.

거래 데이터를 이용한 Box Plot 모델은 특정 사용자의 6개월치 거래 관련 정보를 수집한 후, 해당 데이터를 이용하여 구성하였다.

5.2 실험 결과

SVM 알고리즘을 이용한 사용자 입력 패턴 실험은 SMO 알고리즘에 Polynomial 커널과 RBF(Radial Basis Function) 커널을 사용하여 수집된 1200개의 데이터를 입력하였을 때, 각 파라미터 값의 변화에 따른 결과를 확인하였다. 실험 평가를 위해 10-Folds Cross Validation을 수행하였으며, 실험 결과는 Max FP(False Positive) Rate, 정확도, 모델 생성 시간, 탐지 시간의 4가지 부분을 측정하였다. False Positive는 특정 사용자의 입력 패턴이 아닌 데이터를 사용자의 입력 패턴으로 잘못 구분하는 것을 말한다. 다양한 사용자의 패턴이 한 번에 입력되기 때문에 다양한 FP 값이 나오게 되므로, 이 중 최대(Max) FP Rate에 대해 결과를 확인하였다.

(표 2)는 사용자 입력 패턴 정보에 대한 실험 결과이다.

(표 2) 수집된 사용자 입력 패턴 데이터에 대한 SVM 학습 실험 결과
(Table 2) The experimental result of SVM with collected user input pattern data

입력 값 구분		실험 결과			
커널 함수	함수 파라미터	Max FPR	정확도 (%)	모델 생성 시간(sec)	탐지시간
PolyKernel	Exponent p : 1.0	0.014	98.0392	0.38	< 1 ms
PolyKernel	Exponent p : 1.3	0.013	99.1928	0.36	< 1 ms
PolyKernel	Exponent p : 2.0	0.014	98.0392	0.37	< 1 ms
PolyKernel	Exponent p : 2.5	0.014	98.0392	0.45	< 1 ms
PolyKernel	Exponent p : 3.0	0.015	97.3856	0.38	< 1 ms
PolyKernel	Exponent p : 4.0	0.035	94.7712	0.37	< 1 ms
RBFKernel	Gamma g : 0.005	0.035	92.1569	0.34	< 1 ms
RBFKernel	Gamma g : 0.01	0.021	94.7712	0.48	< 1 ms
RBFKernel	Gamma g : 0.02	0.014	98.6928	0.38	< 1 ms
RBFKernel	Gamma g : 0.03	0.014	98.0392	0.38	< 1 ms
RBFKernel	Gamma g : 0.04	0.014	98.0392	0.37	< 1 ms
RBFKernel	Gamma g : 0.05	0.021	97.3856	0.36	< 1 ms
RBFKernel	Gamma g : 0.06	0.028	96.732	0.38	< 1 ms
RBFKernel	Gamma g : 0.1	0.049	91.5033	0.36	< 1 ms

(표 4) 기존 학습 연구 결과와의 비교
(Table 3) The comparison with previous research

구분	입력값 종류	FPR	학습시간 (sec)	성공률 (%)	실험환경
[9]	12	·	124	98.9	1.3GHz, 3GB
Ours	26	0.013	0.68	99.19	
[10]	3	0.039	·	96	·

커널 함수와 파라미터 값의 변화에 따라 정확도에서 차이를 보였는데 Polynomial 커널을 사용할 경우에는 Exponent 변수 p를 1.3으로 설정하였을 때 약 99.2%의 정확도와 0.013의 Max FP Rate를 보였다. RBF(Radial Basis Function) 커널을 사용할 경우에는 Gamma 변수 g를 0.02로 설정하였을 때 약 98.7%의 정확도와 0.014의 Max FP Rate를 보였다. 모델 생성 시간은 전체적으로 약 0.34~0.48 초 정도로 빠른 생성 시간을 보여주었으며 커널 함수 종류 및 커널 함수 파라미터에 대한 차이는 크지 않았다. 또한 테스트를 위한 데이터를 입력하여 해당 데이터가 어떤 범주로 분류되는지 탐지하는 시간은 모든 실험에서 1ms 이하로 측정되어, 탐지가 실시간으로 가능함을 보였다. 실험 결과 SVM 알고리즘의 Polynomial Kernel(p=1.3)을 사용하였을 경우 기존의 연구[9-10]와 비교하였을 때, 더 빠른 학습 시간과 높은 정확도, 낮은 FPR을 보여줌을 확인할 수 있었다.

(표 3)은 본 논문에서 제안하는 방법과 기존 연구들의 실험 결과를 비교하여 나타낸 것이다. 기존 서호진 등의 연구[9]와의 비교를 위해 해당 연구의 실험 환경인 1.3GHz CPU와 3GB의 메모리를 갖는 컴퓨팅 환경을 구성하고, 해당 환경에서 본 논문의 실험을 진행하였다. 그 결과 기존 연구에 비해 입력 값의 종류는 늘어났지만 기존 124초 대비 0.68초의 빠른 학습시간과 더 높은 분류 성공률을 보였다. Yuxin Meng 등의 연구[10]에서는 학습 시간에 대한 실험을 진행하지 않았기에 실험 컴퓨팅 환경은 실험 결과에 영향을 미치지 않으며 연구 내용에도 실험 환경은 명시되어 있지 않았다. 따라서 본 논문과 입력값의 종류, FPR, 성공률에 대한 결과 수치의 비교만을 수행하였다. 본 논문에서 제안하는 방법은 기존 연구에 비해 입력 값의 개수는 많지만 0.039 대비 0.013의 낮은 FPR과 높은 분류 성공률을 나타내었다.

Box Plot 모델 생성을 위해 수집한 사용자의 6개월치 데이터 100개를 이용하여 가장 먼저 백분위 값을 구하였다. 사용자 거래 패턴을 이용하여 도출한 백분위 값은 (표 4)와 같다. (표 4)의 데이터를 토대로 각각의 패턴 데이터에 대한 백분위 통계를 구하면 각각의 패턴 백분위 평균으로 나올 수 있는 모든 가지 수는 60개가 된다. 해당 데이터들은 사용자의 정상 패턴 자료들이다. 이후 전자 금융 비정상 행위 데이터를 추출하기 위해 전체 즉, 1.0의 값에서 정상 패턴을 뺀 백분위의 계산 값으로 데이터 리스트를 만든다. (표 5)는 비정상 행위를 탐지하기 위한 사용자 패턴 백분위를 오름차순으로 정렬한 데이터 리스트다. (표 5)의 데이터 리스트를 이용해 Box Plot 모델을 구성하기 위한 값을 계산해보면 아래와 같다.

- 전체 데이터 수 : 60개
- 최소값(Min) = 0.413
- 최대값(Max) = 0.940
- 중간값(Median) = 0.763
- 1사분위수(Q1)
  : 계산 결과가 정수가 아니므로 15번째 값과 (15+1)

(표 4) 사용자 거래패턴을 이용하여 도출한 백분위 값
(Table 4) The formulated percentile through user transaction pattern

시간	상대도수	백분위
6시~9시	0	0
9시~12시	16	0.16
12시~15시	45	0.45
15시~18시	28	0.28
18시~21시	3	0.03
21시~24시	8	0.08
24시~3시	0	0
3시~6시	0	0

위치 (IP주소)	상대도수	백분위
191.21.14.8 (집)	45	0.45
172.56.9.46 (회사)	53	0.53
192.55.16.9 (기타)	2	0.02

사전행위	백분위
로그인 성공 확률	0.81
로그인 5회 실패 (오류한도 초과)	0.06
개인정보 변경	0.08
인증방법 변경	0.05

(표 5) 오름차순으로 정렬한 사용자 패턴 백분위
(Table 5) The user pattern percentile in ascending order

0.413	0.447	0.470	0.503	0.510	0.537
0.543	0.547	0.553	0.570	0.587	0.603
0.643	0.657	0.663	0.667	0.670	0.687
0.690	0.697	0.700	0.713	0.720	0.723
0.747	0.753	0.753	0.757	0.760	0.763
0.780	0.787	0.787	0.790	0.790	0.793
0.797	0.797	0.797	0.800	0.803	0.807
0.813	0.820	0.823	0.830	0.837	0.840
0.847	0.853	0.857	0.887	0.893	0.897
0.913	0.920	0.923	0.930	0.937	0.940

번째 값의 평균인 0.665

- 3사분위수(Q1)
  : 계산 결과가 정수가 아니므로 45번째 값과 (45+1) 번째 값의 평균인 0.825
- IQR(Inter-Quartile Range) = 0.16

Box Plot 모델에서는 일반적으로 1.5 * IQR 이상인 Upper fence 위의 값을 이상 데이터로 판단한다. 이를 계산해보면 0.665 + (1.5 * 0.16) = 0.905 이 나온다. 이에 따라 데이터 라인에서 0.905 위의 값은 이상치 값으로 판단할 수 있다. 마찬가지로 Lower Fence인 0.095 이하 값이 들어온 경우 해당 데이터를 이상치로 판단할 수 있다.

6. 결 론

본 논문에서는 모바일 뱅킹 서비스 이용 시 사용자 입력 패턴 및 거래 패턴을 이용하여 이상치를 탐지할 수 있는 방법을 제안하였다. 사용자의 입력 패턴과 거래 패턴 데이터는 개개인을 구별할 수 있는 특징을 갖고 있기 때문에 적절히 이용할 경우 올바른 사용자에 의한 금융 거래와 비정상 거래를 구분하기 위한 자료로 활용할 수 있다.

제안하는 방법은 사용자 입력 패턴을 SVM 알고리즘에 적용하여 입력 패턴을 정상, 비정상, 보류의 3가지 범주로 구분하고, 이 중 보류로 분류된 데이터에 대해서만 거래 패턴을 이용한 탐지 과정을 추가적으로 적용함으로써 전체 탐지 과정의 효율성과 정확도를 높일 수 있다. 또한 거래 패턴은 Box Plot 모델을 이용함으로써 사용자 거래 패턴 분포를 한눈에 쉽게 볼 수 있어 금융권에서 이상치 판단의 임계치를 적정하게 설정할 수 있다는 장점을 갖고 있고, 이상치가 판단된 후에도 정상 패턴에서 벗어난 정도를 정확하게 알 수 있다.

제안하는 방법을 실제 모바일 디바이스에서 수집한 입력 패턴 데이터와 실사용자의 금융 데이터를 사용하여 실험을 진행함으로써, 기존 연구들에 비해 더욱 빠른 학습 시간과 낮은 FPR, 높은 정확도를 나타냄을 확인할 수 있었다. 따라서 작은 오탐으로도 문제가 될 수 있는, 정확성이 매우 중요한 금융 거래에서 적절한 탐지 방법으로 활용할 수 있다. 또한 사용자 패턴을 이용한 탐지 방법은 모바일 단말기 분실에 대한 보안 대응방안으로도 이용될 수 있을 것으로 기대된다.

본 논문에서는 사용자 입력 패턴을 이용한 탐지를 효율적으로 개선하고 입력 패턴 탐지 시 발생할 수 있는 오탐을 거래 패턴을 이용함으로써 줄일 수 있는 방안을 제시하였지만, 두 방안을 연동하는 실험은 진행하지 못하여 탐지 정확도의 정확한 수치는 제시하지 못하였다. 향후에는 실제로 보류 범주로 분류되는 데이터를 추출하여 거래 패턴 데이터를 이용한 탐지의 방법까지 적용하였을 때의 탐지 정확도를 측정할 수 있도록 본 논문에서 제안한 두 가지 방안을 실제 연동하는 탐지 실험을 진행할 예정이다.

참 고 문 헌(Reference)

[1] The bank of Korea, “The current usage state of internet banking service in the first quarter of 2013”, May. 2013

[2] KB Financial Group, “Trend and prospect of the Mobile payment market”, May. 2013

[3] A. Shabtai, Y. Fledel, U. Kanonov, Y. Elovici, S. Dolev, C. Glezer, “Google Android : A Comprehensive Security Assessment”, IEEE Security & Privacy, Vol. 8, Issue 2, pp. 35-44, Mar. 2010

[4] Financial Security Agency, “Major issues and trend report on the financial smart phone”, Jul. 2012

[5] Y.J. maeng, D.O. Shin, S.H. Kim, D.H. Nyang, “A study of the vulnerability analysis and coping method in regard to document falsification of electronic financial transaction”, Review of KIISC, Vol. 20, No. 6, pp 17-27, Dec. 2010.

[6] S.M. Lee, “A recent state about the authentication technology”, The payment settlement and information

technology, Vol. 47, pp. 30-48, Jan. 2012

[7] Telecommunications Technology Association, “Fraud Detection and Response Framework in Electronic Financial Transaction System”, TTAK.KO-12.0178, Dec. 2011

[8] A.D. Luca, A. Hang, F. Brudy, C. Lindner, H. Hussmann, ″Touch me once and I know it’s you!: Implicit Authentication based on Touch Screen Patterns″, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 987-996, May. 2012

[9] H. Seo, H.K. Kim, “Novel Anomaly Detection Method for Proactive Prevention from a Mobile E-finance Accident with User’s Input Pattern Analysis”, Journal of the Korea Institute of Information Security and Cryptology, Vol. 21, No. 4, pp. 47-60, Aug. 2011

[10] Y. Meng, D.S. Wong, R. Schlegel, L. Kwok, “Touch Gestures Based Biometric Authentication Scheme for Touchscreen Mobile Phones”, Information Security and Cryptology, Vol. 7763, pp. 331-350, 2013

[11] H. Kim, S. Lee, “The Phoneme Kernel Technique based on Support Vector Machine for Emotion Classification of Mobile Texts”, Journal of KIISE : Software and Applications, Vol. 40, No. 6, pp. 350-355, Jun. 2013

[12] Y.M. Kim, C.H. Jeong, H.S. Kim, “An Estimation of Risky Module using SVM”, Journal of KIISE : Computing Practices and Letters, Vol. 15, No. 6, pp. 435-439, Jun. 2009

[13] J. Park, K. Kim, I. Han, ″Bankruptcy Prediction using Support Vector Machines″, The journal of MIS research, Vol. 15, No. 2, pp. 51-63, Jun. 2005

[14] S.H. Kim, H.S. Jeon, “Box Plot Algorithm used in Packages”, The Korean Journal of Applied Statistics, Vol. 5, No. 1, pp. 93-102, 1992

[15] R. McGill, J.W. Tukey, W.A. Larsen, “Variations of Box Plots”, The American Statistician, Vol. 32, No. 1, pp. 12-16, Feb. 1978

[16] J.L. Hintze, R.D. Nelson, “Violin Plots: A Box Plot-Density Trace Synergism”, The American Statistician, Vol. 52, No. 2, pp. 181-184, May. 1998

[17] WEKA 3 : Data Mining Software in Java, [Online]. Available : http://www.cs.waikato.ac.nz/ml/weka/

[18] Wikipedia, “Support vector machine”, [Online]. Available : http://en.wikipedia.org/wiki/Support_vector_machine

◐ 저 자 소 개 ◑

민 희 연(Hee Yeon Min)
2012년 성신여자대학교 컴퓨터정보학부 졸업(학사)
2012년~현재 고려대학교 정보보호대학원 석사과정
관심분야 : 전자금융보안, 데이터마이닝, 암호 프로토콜, 패턴인식
E-mail : btn12@naver.com

박 진 형(Jin Hyung Park)
2010년 건국대학교 컴퓨터공학과 졸업(학사)
2012년 고려대학교 정보보호학과 졸업(석사)
2012년~현재 고려대학교 정보보호대학원 박사과정
관심분야 : 암호프로토콜, 스마트폰 보안, 암호 알고리즘, 데이터마이닝
E-mail : ikarisj@naver.com

이 동 훈(Dong Hoon Lee)
1983년 고려대학교 경제학과(학사)
1987년 Oklahoma University 전산학과(석사)
1992년 Oklahoma University 전산학과(박사)
1993년~1997년 고려대학교 전산학과 조교수
1997년~2001년 고려대학교 전산학과 부교수
2001년~현재 고려대학교 정보보호대학원 교수
관심분야 : 암호프로토콜, 암호이론, USN이론, 키 교환, 익명성 연구, PET 기술
E-mail : donghlee@korea.ac.kr

김 인 석(In Seok Kim)
1973년 홍익대학교 전자계산학과(학사)
2003년 동국대학교 정보보호학과(석사)
2008년 고려대학교 정보경영공학과(박사)
2009년~현재 고려대학교 정보보호대학원 전문교수
관심분야 : 전자금융보안, IT 감사, 전자금융법규
E-mail : iskim11@korea.ac.kr