Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.1, Feb. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.1.133

의사결정나무를 이용한 이상금융거래 탐지 정규화 방법에 관한 연구

박 재 훈,[1†] 김 휘 강,[1] 김 은 진[2‡]
[1]고려대학교 정보보호대학원, [2]경기대학교 국제산업정보학과

Effective Normalization Method for Fraud Detection Using a Decision Tree

Jae Hoon Park,[1†] Huy Kang Kim,[1] Eunjin Kim[2‡]
[1]Graduate School of Information Security, Korea University
[2]Department of International Industrial Information, Kyonggi University

요 약

전자금융사기의 고도화와 함께 지능적인 수법들이 동원됨에 따라 전자금융 사용자들의 피해사례가 늘어나고 있다. 이에 대한 대응 방안으로 금융당국은 사용자 구간에 집중된 기존 보안 대책 외에 이의 한계성을 극복하기 위한 이상거래 탐지 시스템의 도입을 확대 권고하고 있다. 이상거래 탐지 시스템은 실시간으로 고객의 거래를 확인하고 이상거래 유무를 판별하여 전자금융 사고를 방지할 수 있도록 하는 시스템으로 거래 정보를 빠르게 분석하여 이상거래를 식별하는 것이 핵심이다. 본 논문에서는 사고 데이터분석을 통해 이상 징후 패턴을 파악하고 탐지 룰을 설정하고, 이렇게 설정된 룰을 기반으로 고객 개인별 거래 패턴과 고객 프로파일을 비교하여 이상거래 여부를 판단하고자 한다. 이때 의사결정나무를 사용하여 탐지 룰을 정규화 하여 효과적으로 이상거래를 탐지 할 수 있도록 하는 방법을 제안하고자 한다. 실증 분석을 위해 국내 모 은행의 전자금융 사고 데이터를 바탕으로 패턴 정보와 고객 프로파일 정보를 도출하였고 이를 통하여 탐지 룰을 설정하였다. 그리고 탐지된 룰을 의사결정나무를 사용하여 정규화 한 결과를 순차적인 탐지 방식과 비교하여 제시된 방안이 효과적임을 확인하였다.

ABSTRACT

Ever sophisticated e-finance fraud techniques have led to an increasing number of reported phishing incidents. Financial authorities, in response, have recommended that we enhance existing Fraud Detection Systems (FDS) of banks and other financial institutions. FDSs are systems designed to prevent e-finance accidents through real-time access and validity checks on client transactions. The effectiveness of an FDS depends largely on how fast it can analyze and detect abnormalities in large amounts of customer transaction data. In this study we detect fraudulent transaction patterns and establish detection rules through e-finance accident data analyses. Abnormalities are flagged by comparing individual client transaction patterns with client profiles, using the ruleset. We propose an effective flagging method that uses decision trees to normalize detection rules. In demonstration, we extracted customer usage patterns, customer profile informations and detection rules from the e-finance accident data of an actual domestic(Korean) bank. We then compared the results of our decision tree-normalized detection rules with the results of a sequential detection and confirmed the efficiency of our methods.

Keywords: Fraud Detection System, Banking System, e-finance accident, Decision Tree, Normalization

접수일(2014년 11월 17일), 수정일(1차: 2014년 12월 29일, 2차: 2015년 1월 21일), 게재확정일(2015년 2월 4일)

† 주저자, tabrisme@gmail.com
‡ 교신저자, ejkim777@kgu.ac.kr(Corresponding author)

I. 서 론

최근 주요 카드사 및 통신사의 개인정보 유출 사건으로 인하여 유출된 개인정보들을 이용한 전자금융사기(보이스피싱, 파밍, 스미싱 등) 및 대출사기가 급속히 증가하고 있다. 금융감독원에 따르면 2014년 상반기 중 피싱사기 피해금액은 886억원 (1.3만건)으로 전년동기 대비 87.7% (건수기준 34.1%) 증가하였고, 사기수법의 교묘성, 피해 인지의 어려움과 함께 피해금 인출은 더욱 빨라져 피해금 환급률은 11.9%로 전년동기(17.1%) 대비 5.2%p 감소하였다[1].

정부는 급격하게 증가하고 있는 전자 금융사기에 대한 피해를 사전에 예방하고자 전자금융사고 시 접근매체의 위조나 변조로 사고가 발생했을 때 고객과실을 입증을 못하면 금융회사나 전자금융업자에 책임을 묻도록 전자금융거래법[2]을 개정 하였다. 이와 더불어 금융위원회는 자율적인 노력을 통하여 전산사고를 방지 할 수 있도록 제도적, 기술적 보안관리 체계 강화에 중점을 두는 금융보안 종합대책[3]을 통해 이용자 보호 강화를 권고 하였다.

그리고 정부는 최근 '전자금융사기 예방서비스' 전면 시행(9.26)[4]을 통해 더욱 적극적으로 전자 금융사기에 대응하려 하고 있다. 하지만 전자금융사기 예방서비스 시행 이후에도 각종 고도화된 금융사기로 인해 피해사례가 계속 늘어나 거래보안수단(공인인증서, 비밀번호, 암호화, 키보드보안 등 다수 보안S/W)만으로는 신종기법을 활용한 사기행위 근절이 어려운 상황이다.

이는 기존의 보안대책이 사용자 구간에만 집중되어 효과적으로 전자금융사기를 예방하기에 한계점을 보여 준다. 이에 따라 부정거래 모니터링, 고객성향, 거래행위분석을 통한 예방의 필요성을 부각시키고 있다. 이에 금융위원회에서 '금융보안 종합대책'을 발표하여 은행용 이상금융거래 탐지 시스템(fraud detection system)구축을 권고 하였고 국회에서 '전기통신금융사기 피해방지 및 피해금 환급에 관한 특별법'(2013.12.31)개정안이 통과되면서 이에 대응하기 위해 각 금융기관들도 이상금융거래 탐지 시스템의 도입을 서두르고 있다.

2014년 10월 15일 보안 업체들은 'FDS (Fraud Detection System) 산업 포럼'을 공식 출범하고 금융회사 및 참여 회원사들의 의견을 반영해 산업 표준 제정, 공통 기술 개발, BMT 평가 대행, 공동 고객 대응 센터 운용, 교육 및 컨설팅 지원, 컨퍼런스 등을 정례적으로 수행하기로 하는 등 업계에서도 본격적인 대응을 하기 시작하였다.

이상금융거래 탐지 시스템은 금융소비자들의 전자금융거래를 실시간으로 모니터링(monitoring)하여, 정책에 위배된 행위가 발생할 경우 즉각 거래를 중단시키고 금융기관과 소비자에게 통보하는데 활용된다. 따라서 이상금융거래 탐지 시스템은 무엇보다 빠르고 정확한 탐지가 매우 중요하다고 할 수 있다.

본 논문에서는 사고데이터 분석을 통해 이상 징후 패턴을 파악하고 이를 고객 개인별 거래패턴과 비교하여 이상거래 여부를 판단한다. 이때 사고사례를 통해 도출한 의사결정나무를 사용하여 탐지 룰을 정규화(normalization)하는 방법을 제안하고자 한다. 실증 분석을 위해 국내 모 은행의 전자금융 사고 데이터를 바탕으로 패턴정보와 고객프로파일 정보를 도출 하였고 이를 통해 제시된 방안이 효과적임을 확인 하였다.

논문의 구성은 먼저 2장에서는 이상금융거래 탐지 시스템의 개념 및 탐지 방법, 구축 현황, 기존 국내외 관련 연구를 살펴본다. 3장에서는 사고 사례를 통하여 사고패턴에 대한 분석을 하고 사고발생고객의 프로파일을 작성하여 의사결정나무를 이용한 이상거래를 탐지한다. 4장에서는 이상금융거래 탐지 분석을 통하여 만들어진 의사결정나무를 바탕으로 정규화하는 방법을 제안하고 사례를 통해서 검증 및 결과를 분석한다. 5장에서는 본 논문에 대한 결론과 연구의 한계를 살펴보고 향후 연구에 대한 방향을 제시하고자 한다.

II. 연구 배경 및 기존 연구

2.1 이상금융거래 탐지 시스템의 개념 및 탐지 방법

이상금융거래 탐지 시스템은 전자금융거래에 사용되는 단말기 정보, 접속정보, 거래내용 등을 종합적으로 분석하여 의심거래를 탐지하고 이상금융거래를 차단하는 시스템을 의미한다[5].

금융보안 연구원에서는 ITU (International Telecommunication Union) 이상금융이상금융거래 탐지 시스템관련 표준(안)[6]에서 제시한 모니터링, 탐지, 차단 3가지 단계를 참고하여 Fig.1.과 같은 구성을 제시하였다.

이의 각 단계별 기능을 살펴보면 다음과 같다. 정보수집 단계에서는 '이용자 매체환경 정보'와 '사고 유형정보'를 수집하고 분석 및 탐지 단계에서는 수집된 정

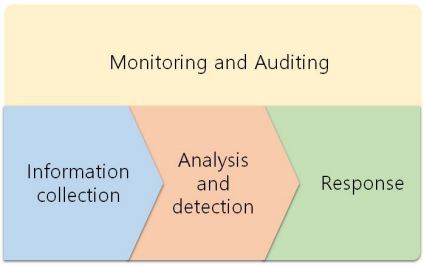

Fig.1. The four configurations of fraud detection system[5]

보를 통해 이상 행위에 대한 정보를 수집한다. 대응 단계에서는 분석된 이상 금융거래 행위에 대한 정상 판별 및 이상거래에 대한 차단 등을 수행, 모니터링 및 감사에서는 수집·분석·대응 등의 종합적인 절차를 통합하여 관리하는 모니터링 기능을 권하고 있다. 본 논문에서는 분석 및 탐지 단계에서 의사결정나무를 통하여 효과적인 이상금융거래 탐지 방법에 대해 연구하고자 한다.

2.2 이상금융거래 탐지 시스템 구축 현황

금융거래의 이상거래탐지는 신용카드사기, 보험사기 방지와 관련된 분야에서 활발히 논의되어 왔고 이에 따라 카드사들을 중심으로 이상거래탐지 시스템이 도입되어 왔다.

신한카드가 지난 1997년에 이상금융거래 탐지 시스템을 처음 도입한 이후 지난 1998년에 비씨카드, 2000년 삼성카드, 2003년 국민카드가 이상금융거래 탐지 시스템을 도입했다. 이후 하나SK카드가 2008년, 현대카드가 2009년에, 지난해 이후에는 우리카드, 롯데카드 등이 모두 이상금융거래 탐지 시스템을 구축해 사실상 전업계 카드사 8곳이 모두 이상금융거래 탐지 시스템을 갖춘 상태다[7]. 이상금융거래 탐지 시스템은 실제로 최근 5년간 8개 카드사에서 약 15만건(148,386건) 가까운 이상거래를 적발해내면서 금융 사고를 방지하는 효과가 있는 것으로 나타나 그 효과성을 입증 했다[8].

해외의 경우를 살펴보면 미국은 대부분의 카드사와 은행이 이상금융거래 탐지 시스템을 구축했음을 알 수 있었다. 대표적으로 홍콩상하이은행(HSBC), 뱅크오브아메리카(BOA), 씨티은행 등이 이상금융거래 탐지 시스템을 구축한 것으로 나타났다. 이는 우리나라 금융감독원에 해당하는 미국 연방금융회사검사위원회(FFIEC)가 미 금융기관들이 이상금융거래 탐지 시스템을 구축하도록 유도한 결과였다[9].

국내도 '금융전산 보안강화 종합대책'으로 인해 은행들도 이상금융거래 탐지 시스템의 도입이 의무화되면서 신한은행과 부산은행이 2013년과 2014년 각각 이상금융거래 탐지 시스템을 구축했고 다수의 은행이 도입을 검토하고 있는 상황이다.

국내의 경우 인터넷뱅킹 서비스에서의 이상거래탐지는 이용 단말이 신용카드의 결제와는 달리 모바일 디바이스, ATM기기에서부터 PC까지 다수의 다기종 매체에서 이용할 수 있는 점, 계좌이체의 특성 상 자신의 구매 이용패턴에 비해 활용가능한 특징값(feature)의 종류가 다양하지 않다는 점에서, 신용카드 결제에서의 이상거래탐지보다 더 난해하고 개인들의 사용 패턴을 학습하기가 쉽지 않다. 게다가 공인인증서와 같은 특수한 환경으로 인해 단순 룰에 의한 탐지만으로는 이상금융 거래를 탐지하기 어렵다. 또한 현재 우리나라 전자금융서비스의 문제점은 공인인증서 탈취가 빈번하게 이뤄지고 있다는 점인데 이상금융거래 탐지 시스템이 공인인증서를 보유하고 있는 사용자를 모두 정상사용자로 취급해서는 안되는 점도 룰 설정이 쉽지 않은 이유이다[10].

또한 탐지룰의 민감도를 높이게 되면 탐지범위(Support)는 넓어지지만, 오탐(False-Positive)이 발생할 경우 민원이 발생하기 때문에 비록 미탐(False-Negative)이 발생하더라도 오탐률이 없는 알고리즘을 개발해야 하는 서비스의 특수성이 있다. 금융 업무의 특성 상 탐지만큼 중요한 것이 고객의 불편을 최소화 하는 것이기 때문이다.

더불어 국내 금융환경에서 이상금융거래 탐지 시스템을 구축할 경우 고려해야 하는 점으로 실시간 탐지에 초점을 맞추어 탐지 알고리즘을 개발해야 한다는 점이다. 외국의 경우 사기행위에 대한 탐지가 실시간으로 이루어져야 하는 제약조건은 거의 없다고 할 수 있다. 외국 은행의 경우 타행으로의 실시간 이체업무를 하는 곳이 많지 않기 때문에 시스템에서 탐지한 후 사람에 의한 분석과 검토 및 고객에 연락하여 확인을 할 수 있는 시간적인 여유가 있는 반면, 국내 은행들의 경우 매 초 타행으로의 대량의 실시간 금액 이체가 일어나기 때문에 이상 금융거래 행위를 탐지하기 위해선 실시간 분석이 무엇보다 중요하다.

이처럼 국내에서 이상금융거래 탐지 시스템을 구축하려면 성능과 확장성이 무엇보다 중요한 사안이다. 다시 말하면 이상금융거래 탐지 시스템으로 인해 본

거래에 영향을 최소화 해야하기 때문이다. 인터넷뱅킹에서 예비거래 시 수집된 로그(log)를 데이터베이스(database)에 저장하고 기존 흐름을 계속 진행하되, 이상금융거래 탐지 시스템은 이를 재빨리 분석해 본 거래가 이뤄지기 전에 차단/추가인증/거래를 할지를 시스템에 알려줘야 한다. 그렇지 않으면 고객 불편으로 인하여 시스템의 효용성이 많이 떨어질 수 밖에 없기 때문이다. 그리고 텔레뱅킹이나 ATM기기 등 채널이 확장하더라도 쉽게 적용하려면 성능이 보장 되어야만 한다[11].

2.3 이상금융거래 탐지 연구 현황

이상금융거래에 대한 탐지 기술은 신용카드, 보험사기와 관련된 분야에서 많은 연구가 이루어져 왔다.

Nagi의 연구[12]를 보게 되면 은행 이상거래(bank fraud)에 대한 이전의 연구들은 상당수가 효과적인 분류(classification)에 대한 알고리즘 연구가 주를 이루고 있다.

Jha의 연구[13]에 따르면 이상거래 탐지에 관한 연구들은 크게 세 가지로 나눌 수 있다. 전통적인 통계적인 기법을 이용한 이상거래탐지(statistic fraud detection), 룰 기반(rule-based)의 탐지방법, 그리고 최근 데이터 마이닝(data mining)에 의한 기법으로 나눌 수 있다. 예를 들면 통계적 기법을 활용한 이상금융거래 탐지는 선형 판별법(linear discriminant)과 로지스틱 회기분석(logistic regression)이 있고 룰 기반의 이상금융거래 탐지는 의사결정나무 기반의 알고리즘을 많이 사용하였다. 시스템 성능이 좋아지고 빅 데이터(big data)에 대한 분석 및 데이터 마이닝 기법들이 발전하면서 연구들은 좀 더 기계학습(machine learning)을 이용한 인공지능 쪽으로 발전하고 있다.

Aihua 등의 연구[14]에서는 의사결정나무, 인공신경망(artificial neural network), 로지스틱 회기분석을 실험을 통하여 가장 효과적인 분류기법에 대한 연구를 진행하였고 의사결정나무보다 인공신경망과 로지스틱 회기분석이 더 뛰어남을 확인 하였고, 나아가 Whitrow 등의 연구[15]는 실험을 통하여 랜덤 포레스트(random forest)가 서포트 백터머신(SVM), 로지스틱 회기분석, K-최근접이웃(KNN) 알고리즘 보다 성능이 더 좋음을 확인하였다.

2.4 기존연구의 검토와 본 연구의 제안방법

2.4.1 기존연구와 현황 검토

앞에서 살펴본 것처럼 기존 연구들은 정상적인 거래와 이상거래를 효과적으로 분류(classification)하는 알고리즘에 대한 연구가 주를 이루고 있다.

하지만 국내 인터넷뱅킹을 비롯한 전자금융거래의 경우 외국의 경우와 직접적으로 비교하기에는 한계가 있고 실시간 이체가 많다는 특수성 또한 고려해야 할 것이다.

김정선의 연구[16]에서 인터넷뱅킹의 금융거래(로그인, 금융정보변경, 전자금융거래)를 기반으로 행위를 분석하고 이상징후인자(IP, MAC Address, 공인인증서 재발급, 수취계좌번호, 이체 금액)을 기반으로 피싱 의심행위 분석을 제안한다. 하지만 이 방식은 거래 행위의 패턴을 벗어나는 이상거래 탐지에는 역부족이고 접속 IP를 우회하거나 공인인증서 재발급과 같은 이상행위를 별도의 시간차를 두고 진행 한다면 탐지하기 어려운 단점이 있다. 그리고 고객정보가 노출되고 보안카드 정보가 노출된 고객의 경우에는 정상적인 거래의 비교군과 대조 했을 때 혐의거래로 판단하기에는 어려움이 있다.

장재환의 연구[17]에서는 통계 분석으로 사고가 난 접속환경 또는 비정상적인 금융거래 유형정보를 통하여 사고 시나리오를 파악하여 탐지에 적용할 수 있는 룰을 생성한다. 그리고 과거 금융거래의 패턴 정보를 바탕으로 가설을 설정하여 의사결정나무를 가지고 발생 확률이 높은 룰을 찾는다. 그리고 이렇게 찾아진 두 가지 룰을 통하여 탐지률을 높이는 방법을 제시 하였다. 하지만 의사결정나무를 탐지 룰을 찾는데만 활용하여 실질적인 이상거래 탐지는 선형적인 탐지에 그치고 있다.

Quah의 연구[18]는 실시간 뱅킹 시스템에 인공신경망보다 진보한 자가조직도(self-organization map) 알고리즘을 적용할 것을 제안하였고, 이는 수행속도가 뛰어나 실시간 학습처리가 가능하고, 통계적 분포도 시간에 따라 변화하여 숨겨진 패턴을 감지 할 수 있는 장점이 있다. 하지만 많은 입력 데이터가 들어오는 전처리 과정이 필요하고 신경망 내부를 추측하기 어려워 결과에 대한 과정 설명이나 추론이 어렵다.

서호진 등 의 연구[19]에서는 모바일 디바이스에서 사용자 입력 패턴(손가락 압력, 면적, 초기 위치, 스크롤 시 속도 및 휠 가속도)을 이용하여 도용 여부를 탐

지하는 방법론으로 BPN 신경망을 이용한 기법을 적용하였으며, 지문정보와 같은 직접적인 생체정보를 이용하는 것이 아니라 간접적인 정보를 이용하는 것만으로도 높은 정확도를 보여주었다.

금융보안연구원의 '이상금융거래 탐지 시스템 기술가이드[5]'는 이상금융거래 구축 시 필요한 기술적인 연구를 하고 있는데 각각 패턴 탐지 모델에 대하여 기술적이고 이론적인 장단점만 설명 할 뿐 실제 구축 시 좀 더 효과적인 방식에 대해서는 비교 분석하고 있지 않다. 금융보안연구원의 기술가이드 특성상 특정 모델에 대해서 권고한다든지 강제화 할 수 없기 때문이다.

2.4.2 본 연구의 제안 방법

이상금융거래 탐지의 최종적인 목표는 효과적인 분석을 통하여 빠르게 실시간으로 이상금융거래를 탐지하고 더불어 고객의 금융거래에 불편을 최소화 하는 것이다. 하지만 기존 이상금융 탐지 알고리즘에 관한 연구들을 살펴보면 단순히 어떤 알고리즘이 효과적이냐에 대한 성능 비교 연구가 다수를 차지한다.

기존의 이상금융거래 탐지 시스템들은 일괄적인 패턴 기반의 의사결정나무 혹은 인공신경망 알고리즘과 탐지 룰에 의한 선형 탐지 방식을 병행해서 사용하는 경우가 많다. 그리고 전처리 단계가 많은 인공신경망과 같은 데이터 마이닝 기법들을 실시간 처리보다는 분석(analysis)에 더 많이 활용되고 있다.

다수의 시스템에 적용되어 있는 탐지 룰에 의한 선형 탐지 방식은 속도가 빠른 장점이 있지만 오탐률이 높은 단점이 존재한다. 또한 선형 탐지 방식은 중복 탐지도 많고 탐지 룰의 개수가 늘어남에 따라 성능 문제에서도 자유로울 수 없다.

그리고 통계에 의한 인공신경망과 의사결정나무를 적용한 시스템도 일괄적인 패턴 기반의 탐지방식을 적용하게 되면 패턴을 벗어난 거래에 대해서 오탐률이 높다.

본 연구에서는 이러한 탐지 방식들의 단점을 보완하고자 사고 데이터 분석을 통해서 이상금융거래 탐지 룰을 설정하고, 고객 개별의 프로파일을 작성하여 이상거래 유무를 판별 할 것이다. 그리고 의사결정나무를 활용하여 탐지 룰 정규화 과정을 거쳐 이상금융거래 탐지의 효율성을 증대시키는 방법을 제안하고자 한다.

III. 이상금융거래 탐지 룰 분석 및 모형화

3.1 사고사례를 통한 패턴 분석

A은행의 2013년 이후 발생한 전자금융사고 실례 500건을 통계에 의한 패턴을 분석하여 Table 1.과 같은 전자금융사고와 관련 있는 이상금융거래 징후를 도출 하였다. 사고시간대는 사고등록시간 30분전 사고등록 이후 6시간으로 지정하고 정상이용기간은 사고발생일 이전 30일로 정의 하였다. 그리고 동일 조건의 정상 거래 비교군 500건을 임의로 추출하여 1000건의 데이터를 통하여 사고와 관계를 분석하였다.

본 연구에서는 1차로 IBM의 SPSS(Statistical Package for the Social Sciences)패키지에 의한 상관관계 분석을 하여 각 변수들과 전자금융거래 시 거래패턴과 사고 발생을 유추하여 각 변수의 영향력을 확인하도록 하고, 통계적으로 분석하기 어려운 항목에 대해서는 별도의 방법을 적용하였다.

우선 사고 발생 내역이 주로 심야 시간대에 발견되

Table 1. Pattern of FDS construction

Classification	Item	Analysis
Transaction period of attacked users	Transaction Time	Midnight transactions (0 a.m.~4 a.m.)
	Initial / Final Transaction	Deviation from normal transaction period
Mediums	New medium	Access with new medium
	Number of mediums	Using 2 or more mediums for attack
	Local	Access from outside of usual local
Daily Transaction to other banks	Daily Transaction frequency	Exceeding daily transaction frequency limit
	Daily Transaction Amount	Exceeding daily transaction Amount limit
Remittance bank	Initial remittance bank	transfer to unprecedented bank (more than 300,000 KRW)
Attacked Savings Account	Withdrawal account balance	Withdrawal minimum balance of Savings Account

는 것에서 착안하여 이체 시간대에 따른 분석을 해 보았다. 정상 이용기간 대비 이체 시간, 이체 빈도, 최초/최종 이체 시간이 평소 이체시간을 벗어난 이체와 사고와의 상관관계를 분석하였고 이용매체를 통한 정보를 가지고 평소 고객의 이용패턴과 다른 정보가 들어왔을 때 상관관계를 분석하였다. 이용매체에 대한 정보는 신규이용 매체를 통한 접속인지, 사고 시간대에 다른 2종 이상의 장비를 통한 접속인지, 평소이용지역을 벗어난 접속인지 확인 하였다. 그리고 고객의 이체 정보에 대한 프로파일을 바탕으로 사고 시간대 일 최대 이체건수 초과 여부와 평소 일일 최대 이체 누계금액을 초과하는지 여부가 사고와 상관관계가 있는지 확인 하였다. 그 다음 송금 은행정보를 바탕으로 이전에 이체 경험이 없는 은행인지 여부를 확인 하였고 특히 공과금 납부 및 단순 지인간의 이체를 제외하기 위해 온라인 소액결제 기준인 30만원 이상 이체 조건을 부여하였다. 그리고 요구불 계좌의 평균 유지 잔액대비 출금을 통하여 상관관계 여부를 확인 하였다. 그렇다면 아래 가설을 통해서 사고와 변수들의 상관관계를 도출해 보도록 하자.

가설. 각 변수들은 사고발생과 상관관계가 있다.

위의 가설을 검증하고자 Pearson 상관계수[1]를 이용 하였고 그 결과는 Table 2.와 같다.

각 변수별로 도출되는 상관계수는 변수와 사고 사이의 상관관계를 보여주는 값이며, 이 값의 범위는 -1과 +1 사이 이다. 괄호 안의 값인 유의 수준(P-value)은 0.05보다 크면 상관관계가 없고, 유의수준이 0.05보다 작으면 상관관계가 있다.

각 변수와 사고발생 여부 간의 관계를 확인한 결과 위의 표와 같은 결과를 확인하였다. 각 변수의 유의수준은 모두 .005 수준보다 유의하였으며, 양의 상관관계를 보여주었다. 이는 각 변수 값이 커질수록 금융이체 시 사고의 발생이 높아지고 있음을 보여주며, 특히 제시된 변수 가운데 최초/최종 이체, 신규이용매체, 일중 최대이체금액, 최초송금은행, 출금계좌잔액 변수는 뚜렷한 상관관계를 보여주었다. 반면 이체시간 변수의 경우 약한 상관관계를 가졌음을 확인 하였다.

그리고 이용매체 수, 이용지역, 일중 최대 이체 건수의 세 가지 변수들은 통계적으로 분석하기에는 기본적으로 표본 자체가 다르기 때문에 다음과 같은 표를 통해 그 차이를 비교하였다.

Table 2. The results of Pearson correlation

Item	Pearson correlation
transaction time	.100 (.002)
initial / final transaction	.199 (.000)
new medium	.239 (.000)
daily transaction amount	.352 (.000)
initial remittance bank	.357 (.000)
withdrawal account balance	.439 (.000)

N=1000
*(P-value)

Table 3. 의 결과를 살펴보면 정상이용기간 사용하는 대부분의 고객은 1개의 매체를 통해서 접속한다. 하지만 사고 시간대 접속매체 수를 살펴보면 복수매체를 통한 동일 시간대 접속이 대조군과 비교해서 급격하게 증가한 것을 알 수 있다. 이는 고객이 아닌 의심거래 행위자의 동시간대 다른 매체를 이용한 접속시도로 간주 할 수 있고 이를 통해서 관계가 있다고 판단할 수 있다.

Table 4. 의 A은행의 접속지역(국가)별 상위 5개

Table 3. Pattern of access by number of mediums

Number of mediums	1	2	3	4	5
normal period	497	3	0	0	0
accident period	280	155	54	9	2

Table 4. Pattern of access from country

Access from country	Korea	China	Japan	USA	Philippines
normal period	497	3	0	0	0
accident period	325	127	23	19	6

1) 피어슨 상관계수(Pearson correlation coefficient)는 $-1 \le \rho \le +1$ 사이의 수를 통하여 두 변수간의 관련성을 구하기 위해 보편적으로 이용되는 개념이다.

국의 접속 통계를 살펴보면 정상이용기간 사용고객의 비율과 사고시간대 접속 지역 비율의 차이를 확인 할 수 있다. 이와 같이 평소와 다른 접속지역의 변경도 사고와 관계가 있는 요소라고 판단할 수 있다.

다음으로 살펴 볼 것은 일중 타행 이체이다. Table 5. 는 사고 고객의 정상이용기간 일일 최대 타행이체 건수를 초과해서 이체가 일어나는 경우와 해당고객의 사고시간대 최대 이체 건수를 초과하여 연속적인 이체가 일어난 수를 나타낸 표이다. 정상이용기간과 비교하여 사고시간대 최대 이체건수 증가가 큰 것을 미루어 특정시간대 타행의 이체건수의 증가는 피해사고와 관련이 있다고 판단 할 수 있다.

Table 5. Pattern of exceeding daily transaction frequency

daily transaction frequency	exceed	not exceed
normal period	68	423
accident period	206	294

3.2 고객 프로파일과 거래를 통한 이상 판별

앞에서 살펴본 9개 항목 중 이체시간(심야시간)은 사고발생과 약한 상관관계를 가졌고 이체와 관련된 변수 중 최초/최종 이체가 상관관계가 더 높으므로 이체 시간 항목을 제외한 8개 항목(최초/최종 이체, 신규이용매체, 피해 시간대 매체 수, 이용국가,일중이체 건수, 일중 최대이체금액, 최초송금은행, 출금계좌잔액)을 이상거래 탐지를 위한 룰로 선정한다.

Table 6.에서 예시는 2014년 8월 중순 사고가 발생하기 전 6개월 동안 전자금융 거래를 바탕으로 A은행에서 발생한 전자금융 사고의 고객 프로파일을 작성한 것이다.

그리고 Fig.2. 는 동일 고객의 사고일 당시 거래내역이다. 이상금융 거래 탐지를 위해서 프로파일과 전자금융 거래를 비교하여 앞에서 설정한 8가지 항목에 대하여 Table 7. 와 같이 평가한다.

Table 6. Profile of banking accident

User ID	AML5**8	Daily maximum withdrawal	600,000 KRW
Normal transaction period	8 am ~ 10 pm	Remittance bank	W bank, S bank
Overseas IP access history	none	Daily transaction frequency	2 cases
Minimum average balance	780,000 KRW	Use medium	1 smart phone

Table 7. Items of FDS rule

No.	Item
1	Did the initial / final transfer time differ from the normal profile?
2	Was the account accessed by a new medium?
3	Was the account accessed by more than one medium during the attack?
4	Which local did the attack originate from?
5	Was the average daily transaction frequency exceeded?
6	Was the average daily transaction amount exceeded?
7	Was the receiving bank have a history usage?
8	Was the account balance below the average daily minimum balance?

인터넷뱅킹은 각 금융회사에서 처리되는 거래절차마다 차이가 있기는 하지만 크게 예비거래와 본거래로 나누어진다. 간단한 예를 들자면 거래가 들어 왔을 때 로그인과 인증서 발급 등의 거래는 사전거래에 속한다. 실시간 이상금융거래 탐지를 위해서 가장 많은 정보를 파악 할 수 있는 구간은 본거래가 실행되기 전 전자서명 전단계일 것이다. 그렇다면 Fig.2.과 같은 거래가 들어 왔을 때 2014년 8월 15일 2:22:24에 들어온 거래를 기준으로 본거래로 정의 하고 이상거래

Customer ID	Transaction date	Transaction time	Transfer	ARS Authentication	Transfer Amount	Balance	using device(OS|H/W)	OS ID	Access IP	Country code	VPN Country	MAC
AML5**8	2014-08-14	23:08:07				820,000	4.1.2.|SHV-E160S	010****7130	218.232.26.173	Domestic		18E2C26FB1C2
AML5**8	2014-08-15	2:19:25				820,000	4.1.1|SHV-E210K	92bcc066bb7c808e	220.117.110.211	Domestic		5CE8EBB357EA
AML5**8	2014-08-15	2:22:24	N		790,000	820,000	4.1.1|SHV-E210K	92bcc066bb7c808e	220.117.110.211	Domestic		5CE8EBB357EA
AML5**8	2014-08-15	2:24:45	N		790,000	820,000	4.1.1|SHV-E210K	92bcc066bb7c808e	220.117.110.211	Domestic		5CE8EBB357EA
AML5**8	2014-08-15	3:21:46				820,000	4.1.1|SHV-E210K	92bcc066bb7c808e	220.117.110.211	Domestic		5CE8EBB357EA
AML5**8	2014-08-15	3:23:49	Y	N	790,000	30,000	4.1.1|SHV-E210K	92bcc066bb7c808e	220.117.110.211	Domestic		5CE8EBB357EA
AML5**8	2014-08-16	19:51:31				30,000	4.1.2.|SHV-E160S	010****7130	223.62.163.78	Domestic		18E2C26FB1C2

Fig. 2. Internet banking log

여부를 탐지해야 한다. 그리하여Table 6. 의 고객 프로파일을 가지고 Table 7.의 항목을 평가하면 Table 8. 와 같은 결과를 얻을 수 있다.

Table 8. The results of compare

Item	Profile	Transaction	Result
Initial / Final Transfer	8 am~10 pm	2 am	Fraudulent
New Medium	4.1.2.|SHV-E160S	4.1.1|SHV-E210K	Fraudulent
Number of Medium during Attack	1	2	Fraudulent
Country	Local Country	Local Country	Legitimate
Daily Transaction Frequency	2	1	Legitimate
Daily Maximum Withdrawal	600,000 KRW	790,000 KRW	Fraudulent
Initial Remittance bank	W Bank, S Bank	W Bank	Legitimate
Balance	780,000 KRW	30,000 KRW	Fraudulent

3.3 의사결정나무를 이용한 이상 금융 거래 탐지

앞에서 확인한 8가지의 이상거래 탐지 룰을 효과적으로 적용하기 위하여 고객의 프로파일과 전자금융거래를 통하여 도출한 Table 8. 의 결과를 바탕으로 실험 데이터를 생성하고 데이터를 바탕으로 의사결정나무를 그린다.

의사결정나무는 어떤 특징을 뿌리마디(root node)로 설정하느냐에 따라 어떠한 분리 기준을 사용하느냐에 따라 여러 가지로 나타낼 수 있는데 본 논문에서는 C4.5알고리즘과 Random Tree알고리즘을 사용하여 의사결정나무를 그리기로 한다. 의사결정나무 구축용 알고리즘의 차이는 어떻게 하면 적절한 크기의 나무를 그리느냐가 핵심인데 본 논문에서는 가중치를 구하는 효율성도 중요하지만 이상거래를 탐지하는 과정을 살펴보는 것도 중요하기 때문에 전체 마디(node)를 살펴 볼 수 있는 알고리즘을 선택한다. 이상금융거래 탐지 시스템에서 담당자는 탐지 룰의 흐름과 상관관계를 파악하고 고객에게 설명 할 수 있어야 하기 때문이다.

의사결정나무를 구성하기 위해 사용된 데이터는 2014년 6월 이후 국내 모 금융사에서 발생한 전자금융사고 30건과 정상거래 70건을 바탕으로 생성하였고, 데이터 마이닝 툴인 WEKA[20]를 이용하여 의사결정나무로 표현하였다. WEKA에서는 C4.5를 조금 개선한 J48을 제공해주고 있는데 본 논문에서는 J48을 사용하여 분석을 하였다. 그리고 앞에서 살펴 본 8가지 이상거래 탐지 룰을 이용하여 Table 9.처럼 마디와 가지를 표기하고 WEKA의 J48을 사용하여 Fig.3. 과 같은 분석결과를 도출 하였다.

결과를 살펴보면 NewDevice를 뿌리마디(root node)로 가지고 4개의 종단마디(leaf node)를 가지고 있으며 의사결정나무의 크기는 7이다. 하지만 전체 데이터 100건 중에 30건이 사고 사례이기 때문에 위의 Fig.3. 에서 yes인 종단마디의 합은 30이 되어야 하나 분류 과정에서 5개의 사고 사례가 부정확하게 분

Table 9. Explanations of node

Item	Node Name	Legitimate	Fraudulent
Initial / Final Transfer	UTime	in	out
New Medium	NewDevice	no	yes
Number of Medium during Attack	Doublelogin	single	double
Country	Ulocation	KOR	CHINA
Daily Transaction Frequency	MaxTrans Cnt	low	high
Daily Maximum Withdrawal	MaxAmount Day	low	high
Initial Remittance bank	FirstBank	no	yes
Balance	Withdraw AcntBal	under	over
Presence of Unusual Transaction	FindFraud	no	yes

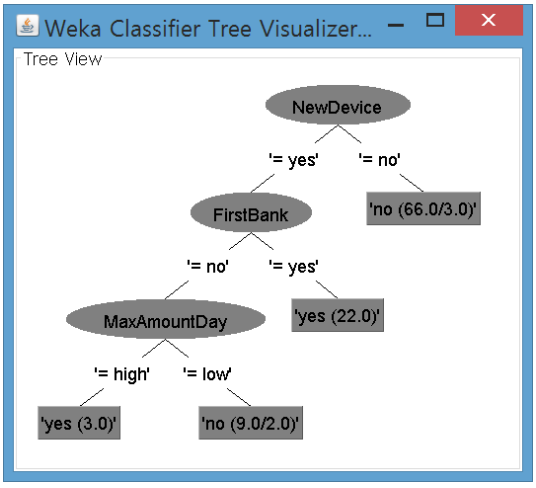

Fig. 3. Decision Tree by J48

류되었음을 알 수 있다.

Fig.4. 는 WEKA의 Random Tree를 사용하여 분석 결과를 도출한 것이다. 결과를 살펴보면 뿌리마디는 FirstBank이고 13개의 종단마디를 가지고 있으며 의사결정나무의 크기는 25이다. 그리고 전체 데이터 100건 중에 30건이 사고 사례가 전부 이기 때문에 위의 Fig.3. 에서 yes인 종단마디의 합은 30이 되어야 하나 Random Tree 또한 분류 과정에서 1개의 사고 사례가 부정확하게 분류 되었다.

앞의 두 가지 WEKA분석 결과를 살펴보면 이상거래 탐지를 위해서 C4.5알고리즘을 사용한 경우가 Random Tree 알고리즘을 사용한 경우보다 나무의 깊이(depth)가 깊어져서 속도 면에서는 불리하지만 정확도는 높아진다.

기존 연구 분석에서 살펴본 바와 같이 Whitrow 등의 연구[15]에서 살펴보면 비교대상의 알고리즘보다 랜덤 포레스트가 더 뛰어난 성능을 보인다는 것을 확인하였다. 그렇다면 의사결정나무를 만들 때 랜덤 포레스트를 사용하여 Random Tree를 만드는 작업을 반복하여 투표로 최종 결과를 도출 한다면 의사결정나무를 사용하는데 더 좋은 성능을 가지고 올 수 있을 것이다.

하지만 본 논문에서는 마디 사이의 관계와 흐름을 파악하고 좀 더 정확한 분류를 통해 정규화를 하는 것이 중요하므로 랜덤 포레스트를 사용하여 나무를 만드는 작업에 대한 설명은 생략하기로 하겠다. 그리고 이렇게 만들어진 Random Tree를 통해서 의사결정나무를 생성하고 이 생성된 의사결정나무를 통하여 이상거래 탐지 룰을 정규화 하도록 한다.

IV. 이상 거래 탐지 룰의 정규화

4.1 의사결정나무를 이용한 정규화

본 논문에서는 고객의 사고사례 패턴을 중심으로탐지 룰을 설정하고 WEKA의 Random Tree 알고리즘을 사용하여 의사결정나무를 만들었다. 그리고 본 장에서는 이렇게 만들어진 의사결정나무를 사용하여

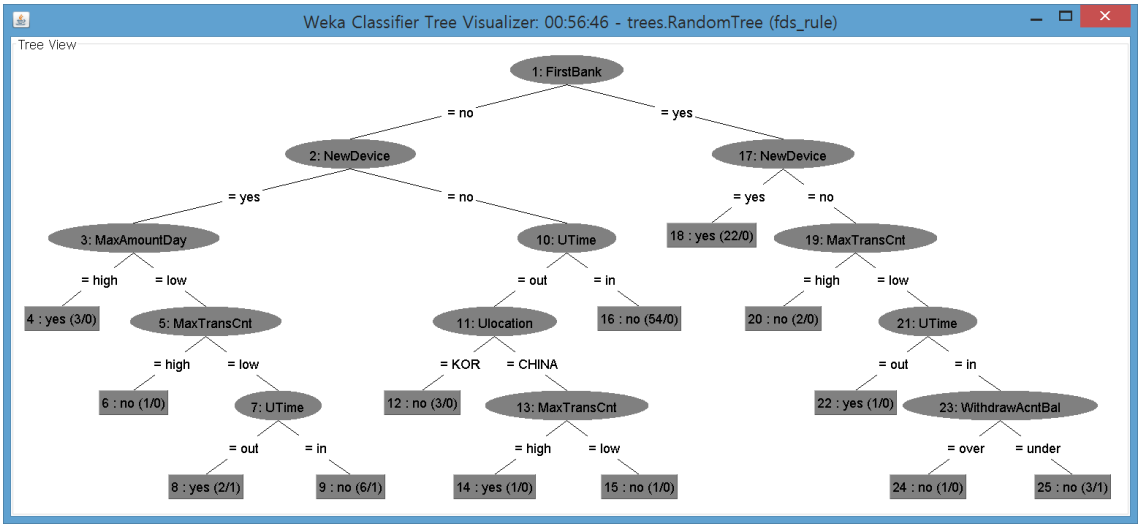

Fig.4. Decision Tree by Random Tree

탐지 룰의 정규화를 제안한다.

기존의 탐지 룰에 의한 이상금융거래 탐지 시스템은 설정된 룰을 테이블 화하여 거래가 들어오면 각각의 탐지 룰을 통하여 Fig.5.와 같이 선형적으로 탐지하였다. 개별의 탐지 룰은 상호간 관계에 관한 정보가 존재하지 않을 뿐만 아니라 개별 탐지로 인하여 중복 탐지에 의한 자원낭비 및 오탐의 확률도 높아지게 된다. 각각의 룰들은 적게는 한 개의 데이터베이스 테이블을 호출하거나 많게는 여러 테이블의 조합에 의해서 이상 유무를 판단하게 되는데 이렇게 룰의 수가 증가하면 할수록 시스템에 부담을 줄 수 밖에 없고 과도한 이상거래 탐지로 인한 고객 민원에 대한 부담도 커지게 된다.

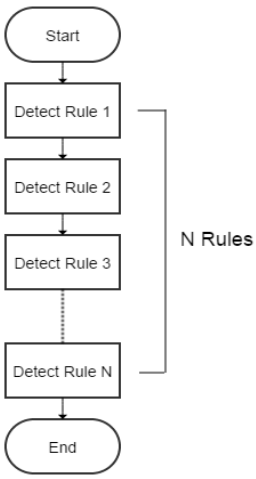

Fig.5. Linear detection model

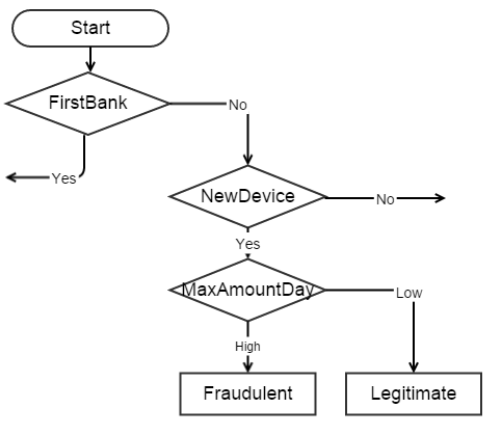

Fig.6. Decision tree detection model

Fig.6. 은 3장에서 만들어진 Fig.4. 의 의사결정나무를 기반으로 Fig.2. 거래의 이상금융거래 여부를 판단하는 프로세스이다.

Fig.6. 의 정규화 된 프로세스를 사용하게 되면 3번의 탐지 룰에 대한 검증만으로 이상금융거래에 대한 탐지 유무를 확인 할 수 있다. 그리고 Fig.6.의 프로세스를 통해 이상거래 여부를 판단하면 아래와 같다.

FirstBank[No] → NewDevice[Yes] → MaxAmountDay[High](profile : 600000 / Input : 790000)

이를 분석해 보면 그리고 Fig.4. 의 전체 의사결정나무에서 위의 판단여부에 영향을 준 마디는 FirstBank, NewDevice, MaxAmountDay 임을 알 수 있고, Table 8. 에서 보게 되면 이상행위로 의심이 되는 판단 결과가 나왔지만 Utime, WithdrawAcntBal 두 가지는 의사결정나무에서 판단의 근거로 사용되지는 않았다.

이를 다시 해석해 보면 Fig.2. 의 거래내역을 가진 고객은 평소에 송금하던 은행에 송금을 했지만 평소에 접속하던 기기가 아닌 새로운 기기를 통해서 평소 통장 잔액이상을 출금하려고 하였다는 것을 알 수 있다. 이처럼 의사결정나무를 사용하면 의사결정의 흐름을 한눈에 파악할 수 있어서 이상거래 탐지로 인한 상담시 담당자가 쉽게 판단의 근거를 이해할 수 있다.

4.2 정규화를 통한 효과

사고 데이터를 바탕으로 구축 된 의사결정나무(Fig.4.)를 통한 탐지와 기존의 선형적 탐지 방식의 이상금융거래 탐지의 차이를 살펴보면 Table 10. 과 같은 결과를 확인할 수 있다. 여기서 사용한 데이터는 3장에서 의사결정나무를 구성하기 위해 사용한 2014년 사례 100건(사고 30건, 정상 70건)을 가지고 수행하였다.

탐지 룰에 의한 선형적 탐지 방식을 적용 했을 때 몇몇 항목은 Table 2. 의 상관관계 분석결과에 의해서 우선순위를 적용할 수도 있으나 모든 탐지 룰이 동일한 방식으로 측정된 결과가 아니므로 서로간의 상관관계의 차이를 적용할 수는 없다. 그리고 탐지 룰 서로간의 관계를 파악할 수가 없기 때문에 각 사례별로 전체 룰에 대해서 탐지를 수행해야 한다.

선형적 탐지 방식으로 100건의 이상거래 사례를 분

Table 10. The comparison of linear detection with tree detection

case	node No	count	linear detection	tree detection	Result
1	4	3	8	3	Fraudulent
2	8	2	8	5	Fraudulent
3	14	1	8	5	Fraudulent
4	18	22	8	2	Fraudulent
5	22	1	8	4	Fraudulent
6	6	1	8	4	Legitimate
7	9	6	8	5	Legitimate
8	12	3	8	4	Legitimate
9	15	1	8	5	Legitimate
10	16	54	8	3	Legitimate
11	20	2	8	3	Legitimate
12	24	1	8	5	Legitimate
13	25	3	8	5	Legitimate

석하려면 각 사례별 8번을 수행해서 800번의 탐지를 수행해야 한다. 반면 의사결정나무에 의한 탐지를 수행하면 100건의 사례를 311번의 탐지를 수행하면 위의 Table 10. 과 같은 결과를 얻을 수 있다. 의사결정나무를 이용하면 가장 많은 탐지 횟수가 다섯 번이고 최소한의 경우에는 두 번의 탐지만으로 이상거래 유무를 판별이 가능하다.

결과적으로 각 탐지 룰마다 데이터베이스 테이블이나 조건 등의 차이가 있겠지만 단순하게 수행 횟수만을 비교 했을 때에도 의사결정나무에 의한 탐지방식이 효과적임을 알 수가 있다.

이에 대한 실제 효과성 검증을 수행하기 위해, 메인프레임(mainframe)환경을 사용하는 A은행 계정계 시스템의 테스트 원시데이터(raw data)를 조합하여 이상거래탐지 데이터베이스 테이블을 테스트 환경에 구축하고, 각 이상거래 탐지 룰에 대한 탐지 쿼리(query)를 간단하게 구현하였다. 원시 데이터 테이블은 5개이고, 탐지 룰에 필요한 컬럼(column)만을 도출하여 이상거래탐지 데이터베이스 테이블은 1개로 구축하였다. 또한, 모든 탐지 룰에 대해 쿼리를 수행하는 선형적 탐지 방식을 처리하는 프로그램 1본과 의사결정 나무를 적용하여 선행된 탐지 룰의 판단 결과에 따라 하위 탐지 룰의 쿼리를 수행하는 프로그램 1본을 작성하여 앞서 구현된 각 이상거래 탐지 룰에 따른 탐지 쿼리를 반영한 후 수행하였다. 테스트 환경은 메인프레임 z10이며, DBMS는 DB2 V10이고, 이상거래 탐지 데이터베이스 테이블의 열(row) 건수는 약 3,200만 건이다.

Table 11. 은 선형적 탐지 방식과 의사결정나무를 적용한 탐지방식의 두 프로그램을 Trace 처리하여, 각각의 탐지 룰에 대한 수행 횟수를 확인한 결과이다.

메인프레임 환경의 z/OS는 프로그램의 CPU 사용 수준을 MIPS(Millions of Instructions Per Second)로 표시한다. MIPS는 일반적으로 CPU용량 및 CPU 처리능력의 측정기준을 의미하며 메인프레임 기반의 응용프로그램 실행과 연관된다. 각각의 쿼리가 수행되는 비용 수준을 나타내는 것을 'path length'라고 하며, 이것은 MIPS로 표시된다[21]. 즉, MIPS 수치가 높은 쿼리 또는 프로그램이 CPU 자원을 더 많이 사용하는 것이다. 따라서, 테스트 대상 거래 100건에 대해 선형적 탐지 방식과 의사결정나무를 적용한 탐지방식의 프로그램을 수행한 평균 path length의 산출이 필요하며, 그 결과는 Table

Table 11. The performed query count of each detection rules about two programs

Item	Linear Detection	Decision Tree Detection
Initial / Final Transfer	100	100
New Medium	100	88
Number of Medium during Attack	100	54
Country	100	30
Daily Transaction Frequency	100	18
Daily Maximum Withdrawal	100	10
Initial Remittance bank	100	7
Balance	100	4

Table 12. The called query count and average 'path length' of two programs about 100 transactions

Linear Detection		Decision Tree Detection	
called query count	average 'path length'	called query count	average 'path length'
800	16.12	311	9.77

12. 와 같다.

Table 12. 에서와 같이, 의사결정나무를 적용하는 경우 선행된 탐지 룰에서 이상거래로 판별되어 하위 탐지 룰의 쿼리를 수행하지 않음으로써 불필요한 탐지 룰에 대한 쿼리 수행을 제거하여 평균 path length가 크게 낮아졌음을 알 수 있다.

일반적으로 특정 시간대의 최대 거래처리량을 TPS(Transactions per second)로 나타내어, 시스템 용량 도입 시 최대 가용수준을 측정하는데, 이는 해당 기업의 거래량에 의존적이므로 이를 낮춰서 시스템 도입 용량에 따른 비용을 절감할 수 없다. 하지만, 의사결정 나무를 적용하는 경우 기존의 선형적 탐지 방식에 비해 'path length'가 낮아짐으로써 이상거래 탐지 시스템 도입 용량을 낮춰 시스템 도입 비용을 아래와 같이 절감할 수 있다.

> 의사결정나무 적용 시 시스템 연간 비용절감액 = 최대 가용 TPS × (선형적 탐지 방식에 따른 프로그램 평균 MIPS - 의사결정나무 적용 방식에 따른 프로그램 평균 MIPS) × 1MIPS당 연간 유지비용

Dr. Howard Rubin은 21개 분야, 133개 회사의 메인프레임 시스템 사용 데이터를 분석한 가트너(Gartner)의 자료를 토대로 해당 회사들이 1백만 달러의 수익을 창출하기 위한 1MIPS 당 평균 유지 비용을 $4,445로 제시한다[22]. 만약, 특정 기업이 1MIPS 당 유지비용은 $4,445 달러이고, 해당 회사의 특정시간대 이상거래탐지를 처리하기 위한 최대 거래량 분석결과가 200TPS라고 한다면, Table.12.의 결과와 같이 테스트를 진행했던 선형적 탐지 방식과 의사결정 나무를 적용한 탐지 방식의 프로그램에 대한 연간 유지비용은 Table.13.과 같다.

Table 13. 에서와 같이 이상거래 탐지 처리를 위해 최대 거래 가용량을 200TPS로 확보해야 하는 기업의 MIPS당 연간 유지비용이 $4,445라고 한다면, 의사결정 나무를 적용한 탐지 방식이 약 564만달러 수준의 시스템 연간 유지비용이 절약됨을 알 수 있다. 즉, 최대 거래 가용량이나 MIPS당 연간 유지비용은 기업이 통제할 수 없기 때문에, 의사결정 나무를 적용하여 이상거래 탐지 거래의 'path length'를 줄이는 것은 비용 절감을 위해 매우 중요한 부분이며, 기업의 성장으로 인해 요구되는 최대 거래 가용량이 높아질수록 비용 절감 수준은 더 크게 나타나므로 의사결정 나무를 적용한 탐지 방식의 적용은 대규모 거래를 처리하는 기업일수록 그 필요성은 매우 크다.

Table 13. The annually system maintenance cost of two detection methods in Mainframe System

Method	TPS	MIPS per Transaction	Cost per 1MIPS	Annually Cost
Linear Detection	200	16.12	$4,445	$14,330,680
Decision Tree Detection	200	9.77	$4,445	$8,685,530

V. 결 론

본 논문에서는 빠르고 효과적인 이상거래 탐지를 위해 의사결정 나무에 의한 정규화를 채택하고 이를 실증적으로 적용해 본 후, 시스템 유지비용 절감 효과를 분석하였다. 국내 모 은행의 실제 사고 데이터 통계를 바탕으로 이상거래의 패턴을 확인하였고, 분석된 패턴을 중심으로 이상금융거래 탐지 룰을 설정하였다. 그리고 최근은행에서 발생한 전자금융사고 사례를 중심으로 사고발생고객의 이용자 정보, 거래정보, 장치정보의 6개월 데이터를 분석하여 고객프로파일을 작성하였다. 그리고 다시 분석된 사고 사례의 이상거래 패턴과 고객 프로파일을 가지고 의사결정나무 알고리즘을 이용하여 탐지 룰을 정규화 하였다.

현재 이상금융거래 탐지에 대한 연구들은 좀 더 다양한 이상거래를 탐지하는 빅 데이터 분석에 집중되어 있지만 실무자의 입장에서는 시스템이 적용되었을 때 빠른 분석과 정확하고 안정적인 운영 환경 또한 중요하다. 이러한 측면에서 의사결정나무는 실무 적용에 용이한 방법으로 복잡한 룰에 의해서 실제 거래가 느려지는 것을 최소화 할 수 있기 때문에 가장 효과적인 방법이라고 볼 수 있다. 또한, 본 논문에서는 대량의 거래를 처리하는 경우 기존의 선형적 탐지 방식에 비해 시스템 연간 유지비용을 크게 줄일 수 있어 비용 효율성 측면의 효과를 확인하였다.

하지만 본 논문의 한계점을 살펴보면 표본 데이터가 적다 단점이 있다. 본 논문의 실험 데이터는 70건의 정상거래와 30건의 사고거래의 표본을 가지고 산출을 한 것이다. 하지만 실제 이상금융거래 탐지 시스템은 더 많은 수의 정상거래 데이터를 처리 할 것이고 사고거래의 비율은 현저히 줄어 들 것이다.

또한, 더 많은 유형의 데이터가 학습될 것이고 의사결정나무는 좀 더 복잡해질 수 밖에 없다. 게다가 전자금융사기 방식은 하루가 다르게 발전하고 있기 때문에 좀 더 다양한 데이터와 실험을 통해서 탐지 방식을 발전시킬 필요성이 있다. 그리고 좀 더 효율적인 시스템을 구축 하려면 기계학습에 의해서 의사결정나무를 매일 혹은 정기적으로 재구성하여 좀 더 정교하게 발전시켜 나아가야 할 것이다.

References

[1] The Phishing Fraud Occurrence and Damage Reimbursement Status, the Financial Supervisory Service, May. 2014.

[2] The Electronic Financial Transaction Act Article 9, May. 2013.

[3] Financial Security Comprehensive Plan, the Financial Services Commission, Nov. 2013.

[4] The Electronic Banking Fraud Prevention Service Enforcement, Sep. 2014.

[5] Financial Security Researchers, Fraud Detction System Technical Guide , Aug. 2014.

[6] The 3rd revised text for X.sap-7: Technical capabilities of fraud detection and response for services with high assurance level requirements (TD0250), Apr. 2013.

[7] Busan Financial News, http://busan.fnnews.com/news/201410091642138054

[8] Herald Business, http://biz.heraldcorp.com/view.php?ud=20141007000984

[9] ChosunBiz, http://biz.chosun.com/site/data/html_dir/2014/05/24/2014052401547.html

[10] Digital Daily, http://www.ddaily.co.kr/news/article.html?no=111498

[11] Digital Daily, http://www.ddaily.co.kr/cloud/news/article.html?no=122299

[12] E. Ngai, et al. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decision Support Systems vol. 50, no. 3, pp. 559-569, 2011.

[13] S. Jha, et al. "Employing transaction aggregation strategy to detect credit card fraud." Expert systems with applications vol. 39, no. 16, pp. 12650-12657, May. 2012.

[14] Shen Aihua, "Application of Classification Models on Credit Card Fraud Detection," International Conference on Service Systems and Service Management, Jun. 2007.

[15] C. Whitrow, et al. "Transaction aggregation as a strategy for credit card fraud detection," Data Mining and Knowledge Discovery vol. 18, no. 1, pp. 30-55, 2009.

[16] J.S. Kim, "Trading for over phishing detection assay fraud prevention," Information Security Journal, 23(6), pp. 41-48, Dec. 2013.

[17] J.H. Jang, "A Study on Fraud Detection Technique using Financial Transaction Analysis in Internet Banking," Chung-ang Univ., Feb. 2012.

[18] J.T.S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert systems with applications vol. 35, no. 4, pp. 1721-1732, Dec. 2007.

[19] Hojin Seo, Eunjin Kim, and Huy Kang Kim. "A novel biometric identification based on a users input pattern analysis for intelligent mobile devices," International Journal of Advanced Robotic Systems, Sep. 2012.

[20] WEKA 3.6.11 : Data Mining Software in Java,[Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

[21] Path lengths and CPU time, http://www-01.ibm.com/support/knowledgecenter/#!/SSAUTT_4.1.0/com.ibm.db2tools.anl

41.doc.ug/topics/anluc_pathcputime.htm
〔22〕 Dr. Howard Rubin, "Economics of Computing -The Internal Combustion Mainframe 〔Expanded Version〕," Technology Economics, pp.1-2, 2010.

〈저자소개〉

박 재 훈 (Jae Hoon Park) 정회원
2007년 8월: 홍익대학교 컴퓨터공학과 학사 졸업
2007년 6월~현재: KB국민은행 재직 중
2013년 5월~현재: 고려대학교 정보보호대학원 석사과정
〈관심분야〉 금융보안, 침입탐지, 데이터 마이닝

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학 학사
2000년 2월: KAIST 산업공학과 석사
2004년 5월~2010년 2월: NC소프트 정보보안실장, Technical Director
2009년 2월: KAIST 산업 및 시스템공학과 박사
2010년 3월~현재: 고려대학교 정보보대학원 조교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌직, 침입탐지시스템, 봇넷탐지

김 은 진 (Eunjin Kim) 정회원
1999년 2월: KAIST 산업경영학과 졸업
2001년 2월: KAIST 경영공학과 석사 졸업
2007년 8월: KAIST 경영공학과 박사 졸업
2008년 9월~현재: 경기대학교 국제산업정보학과 조교수
〈관심분야〉 경영정보시스템, 보안경제학