Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.6, Dec. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.6.1525

결제로그 분석 및 데이터 마이닝을 이용한 이상거래 탐지 연구 조사*

정 성 훈,[1†] 김 하 나,[1] 신 영 상,[2] 이 태 진,[2] 김 휘 강[1‡]
[1]고려대학교 정보보호대학원, [2]한국인터넷진흥원

A Survey of Fraud Detection Research based on Transaction Analysis and Data Mining Technique*

Seong Hoon Jeong,[1†] Hana Kim,[1] Youngsang Shin,[2]
Taejin Lee,[2] Huy Kang Kim[1‡]
[1]Graduate School of Information Security, Korea University,
[2]Korea Internet & Security Agency

요 약

금융 산업과 IT 기술의 결합으로 지불 방법이 간편화됨에 따라 소비자의 지불 수단이 현금 결제에서 신용카드, 모바일 소액결제, 앱카드 등을 이용한 전자결제로 변화되고 있다. 이에 전자금융결제를 악용하여 이상거래를 시도하는 사례가 증가하는 추세로, 금융사는 이상거래로부터 소비자를 보호하기 위해 FDS(Fraud Detection System)를 구축하고 있다. 이상거래 탐지 시스템은 실시간으로 이용자 정보와 결제 정보를 분석하여 높은 정확도로 이상거래를 식별하는 것이 목표이다. 본 연구에서는 결제로그 분석 및 데이터 마이닝을 이용한 이상거래 탐지 연구 동향을 조사하였으며, 이상거래 탐지에 사용된 데이터 마이닝 알고리즘을 정리하고 이상거래 탐지 연구를 사용된 데이터 셋, 알고리즘, 연구 관점으로 분류하였다.

ABSTRACT

Due to a rapid advancement in the electronic commerce technology, the payment method varies from cash to electronic settlement such as credit card, mobile payment and mobile application card. Therefore, financial fraud is increasing notably for a purpose of personal gain. In response, financial companies are building the FDS (Fraud Detection System) to protect consumers from fraudulent transactions. The one of the goals of FDS is identifying the fraudulent transaction with high accuracy by analyzing transaction data and personal information in real-time. Data mining techniques are providing great aid in financial accounting fraud detection, so it have been applied most extensively to provide primary solutions to the problems. In this paper, we try to provide an overview of the research on data mining based fraud detection. Also, we classify researches under few criteria such as data set, data mining algorithm and viewpoint of research.

Keywords: Survey, Categorization, Financial Fraud, Fraud Detection, Data Mining, Credit Card

Received(11. 17. 2015), Modified(12. 10. 2015), Accepted(12. 11. 2015)

* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.R0132-15-1004, 모바일 결제사기 공격 역추적 및 피해 방지를 위한 프로파일링 기반 통합대응 기술 개발)

† 주저자, seonghoon@korea.ac.kr
‡ 교신저자, cenda@korea.ac.kr(Corresponding author)

I. 서 론

국내 및 해외에서 금융 산업과 IT 기술의 발전에 따라 두 기술을 융합한 전자금융거래 시장의 규모가 빠르게 증가하고 있다. 온라인 거래뿐만 아니라 오프라인에서도 신용카드를 통한 지불 비율이 현금을 넘어서고 있으며, 스마트폰과 핀테크의 대중화로 인해 모바일 결제 및 앱카드 사용 비중 또한 커지고 있다.

한국은행의 2014년 지급 수단 이용행태에 대한 조사결과에 의하면 소비자는 지급 수단 선택 시 편리성을 가장 중요한 요소로 인식하고 있으며 가장 편리한 지급 수단이 신용카드라고 응답했다[1]. Fig. 1. 은 여러 국가에서 지급 수단별 거래 금액의 규모를 나타낸 것이다. 전체 거래 금액 중 신용 카드를 이용한 결제 금액이 한국 71%, 캐나다 71%, 네덜란드 64%, 미국 55%, 호주 50%, 프랑스 46% 등 신용카드를 이용한 거래가 가장 높은 비율을 차지하는 것으로 나타났다[1].

전자금융거래 기술의 발달 및 간편결제의 등장으로 인증 절차가 간소화됨에 따라 사기결제 및 부정결제의 위험이 증가하고 있다. 여신금융연구소는 국내에서 한 해 평균 200억 원대 규모의 신용카드 부정사용이 발생하고 있으며 2012년에는 보이스피싱으로 인하여 300억 원대 규모의 신용카드 부정사용이 발생하였다고 밝혔다[2].

신용카드 부정사용의 유형으로는 도난 및 분실, 명의도용, 신규카드 미수령, 카드위변조, 카드정보 도용 등이 있다. 특히 피싱(Phishing), 파밍(Pharming), 스미싱(Smishing) 뿐만 아니라 카드 정보 유출 사고로 인해 카드 정보가 유출되면서 카드정보 도용 사고가 많이 발생되고 있는 추세이다. 이에 정부는 '전자금융사기 예방서비스'를 시행하여 전자 금융 사기에 대응하고자 하였으나 지능화된 공격 수단으로 인한 피해 사례가 지속적으로 발생하면서

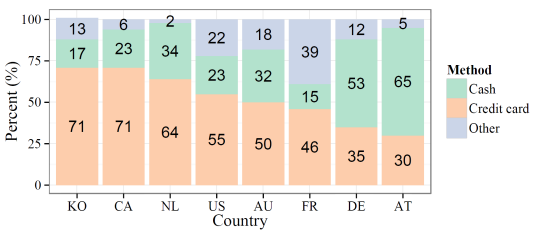

Fig. 1. Comparison between countries' payment method

기존의 키보드 보안, 공인 인증서, 추가 비밀번호 설정만으로는 금융사기에 대응하기 어려운 상황이다[3].

이상거래 탐지 시스템은 실시간으로 이용자의 데이터 및 결제 데이터를 분석하여 평소 패턴과 다른 거래임이 탐지될 경우 금융기관과 이용자에게 탐지 사실을 알리고 더 나아가 임의로 거래를 중단시키는 데 활용된다. 따라서 이상거래 탐지 시스템은 빠르고 정확한 탐지가 중요하고 알고리즘을 향상시키기 위한 연구가 필요하다.

본 연구에서는 전자결제 로그 분석과 데이터 마이닝(data mining) 기법을 이용한 이상거래 탐지 기법에 대해 조사하였다. 각 연구별로 이상거래 탐지를 수행하기 위해 사용된 데이터 셋을 은행 거래, 신용카드, 모바일 거래 등으로 분류하고 비교하였다. 사용된 데이터 마이닝 알고리즘의 특징과 장단점을 정리하고 각 연구를 사용된 알고리즘에 따라 분류하였다. 또한 정확도, 실시간 탐지 가능 여부 등 각 연구별 특징에 따라 분류하여 이상거래 탐지 연구의 최신 동향을 제시한다.

II. 연구 배경 및 동향

핀테크(fintech)는 금융(financial)과 기술(technique)의 합성어로 모바일 결제 및 송금, 개인자산관리, 크라우드펀딩 등 정보기술(IT)을 기반으로 한 새로운 형태의 금융 기술을 의미한다. 국내에서는 2014년 '천송이 코트' 사건 이후 본격적인 이슈가 되었으며 금융관리, 금융데이터 분석, 금융 소프트웨어, 플랫폼의 4가지 영역으로 분류된다. 핀테크의 목표는 사용자들이 편리하게 금융서비스에 접근하여 안전하게 지불할 수 있도록 하는 것이다.

간편결제는 온라인에서 금융거래 시 지급결제에 필요한 개인정보와 신용정보를 전달하는 과정을 단순화시키는 서비스를 말한다. 국내에서는 신용카드사, 오픈 마켓, PG(Payment Gateway)사에서 소액결제에 대해 간편결제 서비스를 제공 중이며 공인인증서 사용 의무화 폐지와 간편결제 도입을 위한 규제 완화 이후 다양한 기업에서 서비스 확대가 추진 중이며 SK플래닛의 시럽페이, G마켓의 스마일페이, 삼성의 삼성페이 등이 있다. 핀테크와 간편결제의 활성화로 고수준의 보안시스템이 요구되고 있으며 금융당국도 2014년 이후 제 1금융권 및 제 2금융권을 중심으로 FDS 구축을 의무화하였다.

핀테크는 각 나라의 소비 패턴과 정책에 따라 적

용 및 응용 범위가 다르다. 한국의 경우 신용카드, 인터넷 뱅킹 등 결제 및 송금이 실시간으로 처리되는 것에 중점을 두고 있고, 중국은 신용카드 시스템, 지급결제 인프라 등이 미비하여 결제대금 예치 방식 서비스를 개발하였다. 미국은 페이팔 등 결제정보가 노출되지 않으면서 신속하고 간편한 결제서비스를 중심으로 핀테크가 활성화 되었으며 금융 거래 시 사전 탐지보다는 사후 탐지에 초점을 맞춰 보안을 강화하는 방식을 사용하고 있다.

비대면 거래가 세계적으로 활성화되면서 이상거래 탐지 연구의 필요성이 대두되고 있다. 여러 시스템에 탑재 가능한 룰 기반(rule-based) 탐지 알고리즘은 속도가 빠른 장점이 있지만 알려지지 않은 이상거래를 탐지하지 못하고 각 나라 별 정책이나 결제 패턴이 다르기 때문에 각 시스템 별로 이상적인 룰(rule)을 결정해야 한다는 단점이 있다. 반면 데이터 마이닝을 이용한 이상 거래 탐지는 대량의 결제로그로부터 유의미한 규칙을 찾을 수 있고 시스템이 고성능화됨에 따라 계산 량이 많이 필요한 단점을 극복할 수 있으므로 현재 이를 이용한 많은 연구가 진행중이다.

Nagi 등은 금융사기 탐지를 위한 데이터 마이닝 기법 적용에 대한 문헌을 고찰하였다. 1997년부터 2008년 사이에 발표된 49개 저널의 논문을 네 가지의 사기 범주와 여섯 가지의 데이터 마이닝 기법으로 분석하고 분류하였다[4].

Sanjeev 등은 실제 신용카드 사기거래 데이터를 통해 사기 유형 분류 및 사기 거래 유형의 국가별 발생 빈도 및 금액을 분석 및 분류하고 박스 플롯(box plot)을 이용하여 분포를 시각적으로 표현한 논문이다. 또한 신용카드사기 행위는 크게 신청 사기 범주와 행동 사기 범주로 분류될 수 있다고 제안하였다[5].

Michael 등은 사기 거래 탐지와 관련된 기존 연구의 포괄적인 조사를 제공하고 특히 사기 거래를 실시간으로 수행하기 위해 시그니처(signature) 기반의 사기 거래 탐지 관련 연구와 사기 거래 탐지 모델을 제시하였다[6].

III. 데이터 마이닝 알고리즘

빅 데이터 분석 기법의 발전과 함께 시스템의 성능이 좋아지면서 이상거래 탐지 시 기계학습(machine learning)을 적용한 연구가 진행되어 왔다. 다양한 데이터 마이닝 알고리즘이 제공되고 있으며, 구축하려는 FDS의 특성과 목표를 고려하여 적절한 데이터 마이닝 알고리즘을 채택해야 한다. 실시간 분류, 예측 및 분류 성능, 입력 및 출력 데이터의 종류 등 이상거래 탐지 시스템의 목표에 따라 여러 알고리즘을 고려할 수 있다.

데이터 마이닝 적용 분야는 분류(classification), 군집화(clustering), 예측(prediction), 이상치 검출(outlier detection), 회귀(regression), 시각화(visualization) 등 6가지 기법으로 분류할 수 있다. 분류 기법은 기존의 데이터와 정답지를 학습하여 모델을 세우고, 정답지가 없는 데이터를 모델에 따라 분류할 때 주로 사용된다. 군집화 기법은 각 데이터의 거리나 속성을 정답지 없이 비교 분석하여 비슷한 데이터들을 하나의 집합으로 정의하는데 사용된다. 예측 기법은 데이터 흐름이나 정렬 규칙 등을 분석하여 다음에 나타날 데이터 규칙을 추측하는데 사용된다. 이상치 검출 기법은 군집화 기법과 유사하게 각 데이터의 거리를 측정하지만 전체 데이터로부터 동떨어진 특정 데이터와 규칙을 찾는데 사용된다. 회귀 기법은 각 데이터를 구성하는 독립 변수 사이의 상관관계를 통계적 방법으로 규명하고 수학적 규칙으로 나타낼 수 있어 이상거래 규칙을 세울 때 많이 사용된다. 시각화 기법은 각 데이터를 최적화된 방법으로 표현하여 데이터 분석가로 하여금 데이터 셋의 특성을 이해하도록 하는데 도움을 준다.

Table 1.은 이상거래 및 사기 행위 탐지에 관한 연구에서 사용된 데이터 마이닝 알고리즘 및 방법론을 나타낸 것이다. 의사결정 트리(decision tree)는 의사결정 규칙(decision rule)을 나무구조로 도표화하여 분류를 수행하는 분석 방법이다. 의사결정 트리는 주어진 데이터를 분류하는 데 주로 사용된다. 또한 분류 과정이 나무구조에 의한 추론규칙에 의해 표현되기 때문에 신경망, 판별분석, 회귀분석 등 다른 방법에 비하여 연구자가 그 과정을 쉽게 이해하고 설명할 수 있다.

랜덤 포레스트(random forest)는 다수의 결정 이진(binary) 트리를 앙상블 형태로 결합한 것으로, 학습 과정에서 랜덤 포레스트는 랜덤한 수의 노드를 생성시키고 각 트리의 노드마다 최적의 판별식과 임계값을 결정한다. 이렇게 생성된 임의의 수의 이진 결정트리들은 앙상블로 결합되어 패턴 분류의 일반화(generalization)율을 높이는 역할을 한다. 특히

Table 1. Data mining algorithms

Algorithm	Description
Decision Tree	- It separates out the complex problem into many simple ones and resolves the sub-problems. - To provide many advantages such as high flexibility and good haleness - A non-parameter method without any assumption for the data distribution
Random Forest	- Ensemble learning technique using decision tree - To provide high accuracy and complement susceptibility to specific data
Logistic Regression	- A regression model to cover the case of binary dependent variables - It measures the relationship among the categorical dependent variables by estimating probabilities using a logistic function.
SVM(Support Vector Machine)	- A supervised learning model with associated learning algorithm to analyze data and to recognize patterns - It represents data as points in space so that the data are divided by a clear gap into separate categories.
ANN(Artificial Neural Network)	- A mathematical model implementing the biological neural networks - It uses to solve a wide variety of tasks that are hard to solve ordinary rule-based programming.
SOM(Self-Organization Map)	- An unsupervised learning model that is useful for visualizing low-dimensional views of high-dimensional data
Naive Bayes	- A simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions among the features
Bayesian Network	- A probabilistic graphical model that represents a set of random variables and their conditional dependencies
Association Rule	- It identifies strong rules discovered in databases using different measures of interestingness.
kNN(k-Nearest Neighbor)	- It assumes that all instances correspond to points in an n-dimensional Euclidean space. - Classification done by comparing feature vectors of the different points
Discriminant Analysis	- It is used to analyze relationships between a non-metric dependent variable and metric or dichotomous independent variables. - It attempts to use the independent variables to distinguish among the groups or categories of the dependent variable.
HMM(Hidden Markov Model)	- It build models representing the hidden states of a process or structure using only observations. - The sequence of tokens generated by an HMM gives some information about the sequence of states.
AIRS(Artificial Immune Recognition System)	- It is based on the human body immune system that is made up of various cells having different functions. - To generates detectors for all class in the dataset

불균형 데이터(imbalanced data)에서 다른 학습 알고리즘보다 좋은 성능을 보이며, 단일 의사결정 트리 모델이 특정 학습 데이터에 민감하거나 불안정함을 보완한다.

로지스틱 회귀분석(Logistic regression)은 분석하고자 하는 대상이 두 집단 혹은 그 이상의 집단으로 나누어진 경우 개별 관측치들이 어느 집단으로 분류될 수 있는가를 분석하고 예측하는 모형을 개발하는데 사용되는 대표적인 통계 알고리즘이다. 종속변수가 바이너리(binary)인 경우 좋은 성능을 보이므로 사기 유무를 결정하는데 적합하다. 분석 목적이나 절차에 있어서 일반 회귀분석과 유사하나 종속 변수가 명목척도로 측정된 범주형 질적 변수인 경우에 사용한다는 점에서 일반 회귀분석과 차이가 있다. 예측변수에 범주형 변수를 투입할 수 있다는 장점이 있다.

SVM(Support Vector Machine)은 지도 학습

에서 사용되는 방법으로, 주어진 자료를 분리하는 초평면 중에서 자료들과 가장 거리가 먼 초평면을 찾는데 사용되는 알고리즘을 제공하며 너저분한 패턴들의 분포들을 가장 잘 분류할 수 있는 기준선을 제공한다. 마진(margin)을 최대화시킴으로써 과적합(overfitting)을 최소화 시켜준다. 커널 함수(kernel function)을 사용하여 비선형 문제를 새로운 공간에서 선형 판별을 수행하며 커널 함수에 따라 계산의 복잡도가 커져 시간이 오래 걸리는 단점이 있다.

인공신경망(Artificial neural network)은 뇌 기능의 특징 몇 가지를 컴퓨터 시뮬레이션으로 표현하는 것을 목표로 하는 수학 모델이다. 복잡하거나 해답이 알려지지 않은 문제를 스스로 학습하여 만족할 만한 해를 제공하지만 문제 해결 과정을 알기 어렵다는 단점이 있다.

SOM(Self-Organization Map)은 주어진 입력 패턴에 대하여 정확한 해답을 미리 주지 않고 자기 스스로 학습할 수 있는 능력을 제공한다. SOM은 전통적인 선형 통계적 방법에서는 할 수 없는 일차 또는 이차 데이터를 쉽게 이해할 수 있는 비선형 관계로 이루어진 복잡한 다차원 입력 데이터를 맵핑할 수 있고, 독립적인 성분 분석과 유사한 데이터를 분석할 수 있다. 또한 SOM은 비선형 관계로 이루어진 복잡한 다차원 입력 데이터를 맵핑할 수 있으므로 숨겨진 패턴을 찾아 새로운 패턴을 제시할 수 있다.

Naive Bayes는 Bayes 이론을 이용한 확률 기반의 분류 방법으로 대량의 데이터에 주로 사용된다. 전체적인 분류 정확도를 최대화하거나 특별히 관심있는 클래스에 속하는 레코드를 식별할 때 사용할 수 있다. Naive Bayes를 구현하기 위한 구상 및 가정이 단순함에도 불구하고 높은 성능을 제공하며, 분류에 필수적인 파라미터를 추정하는데 필요한 훈련 데이터의 양이 적어 속도가 빠르다는 장점이 있다.

베이지안 네트워크(Bayesian network)는 확률변수의 집합과 조건부 독립성을 그래프 구조로 나타내는 확률 그래프 모델이다. 확률 추론을 하는데 유용한 구조를 가지고 있어 불확실한 상황 하에서 결정 문제를 나타낼 수 있고 해결할 수 있다는 장점이 있다.

연관성 분석(Association rule) 알고리즘은 확률변수의 집합과 조건부 독립성을 그래프 구조로 나타내는 확률 그래프 모델이다. 확률 추론을 하는데 유용한 구조를 가지고 있어 불확실한 상황 하에서 결정 문제를 나타낼 수 있고 해결할 수 있는 장점이 있다.

kNN(k-Nearest Neighbor) 알고리즘은 분류되지 않은 개체를 이미 분류된 개체 중 가장 비슷한 속성을 가진 집합으로 할당하는 알고리즘으로, 기존 데이터 중 가장 유사한 k개의 데이터를 이용해서 값을 예측한다. 학습 과정이 빠르며 분류와 숫자 예측에 모두 적용 가능하지만 각각의 데이터에 대한 거리를 구해야 하므로 분류 시간이 오래 걸린다는 단점이 있다.

판별 분석(Discriminant analysis)은 측정변수에 의한 개체를 분류하고 분류되어 있는 집단 간의 차이를 의미 있게 설명해 줄 수 있는 독립변수를 탐색하는 알고리즘이다. 독립변수는 간격척도 또는 비율척도로 측정되었으나 종속변수는 범주척도로 측정된 경우 독립변수와 종속변수의 관계를 조사하는 분석 방법으로 마케팅 분석에 유용하게 이용되는 방법이다.

HMM(Hidden Markov Model)은 확률 통계 모델로 상태 전이 확률과 상태에서 관측 객체가 나타날 확률을 계산하고, 시퀀스 데이터를 통해 새로운 데이터에 대한 생성 확률을 추정하는 알고리즘이다. 순서를 가지는 시퀀스(sequence) 데이터를 사용해 시간적 개념을 포함시킬 수 있으며, 구조가 단순해 데이터에 적합한 모델로 변형하기 쉬운 장점이 있다.

AIRS(Artificial Immune Recognition System)는 척추동물 면역 시스템을 구현한 AIS(Artificial Immune System)를 분류에 적용한 모델이다. AIRS는 다양한 파라미터가 존재하며 실험에 가장 적합한 파라미터를 선택할 경우 다른 분류 알고리즘에 비해 높은 성능을 제공한다. 또한 각 클래스에 대한 일반화 속도가 빠른 장점이 있다.

IV. 결제 로그 분석 및 이상거래 탐지

전자금융거래 환경에서는 다양한 이상 거래가 발생할 수 있다. 따라서 이상거래 탐지 시스템은 거래 시스템의 특징과 환경에 따라 다른 알고리즘으로 구성되어야 한다. 본 절에서는 데이터 마이닝을 이용한 이상거래 탐지 연구를 여러 가지 관점으로 분류한다. Table 2.는 전자금융거래를 위한 이상거래 탐지 및 대응에 대한 연구들을 다양한 관점으로 정리한 것이다. 이상거래 탐지 및 대응에 대한 연구를 사용된 데이터 셋, 알고리즘, 연구 목적 등으로 분류하였다. 또한 각 세부 분류별로 속한 연구들 중에서 해당 분류를 대표할 수 있는 연구를 선정하고 이를 설명하였다.

Table 2. Research classification of fraud detection in financial transactions

Classification	Subcategories	
Data set	- Credit card transaction - Banking transaction	- Telecommunication - Simulation
Data mining algorithm	- Decision Tree - Random Forest - Logistic Regression - SVM(Support Vector Machine) - ANN(Artificial Neural Network) - SOM(Self-Organization Map) - Naive Bayes	- Bayesian Network - Association Rule - kNN (k-Nearest Neighbor) - Discriminant Analysis - HMM(Hidden Markov Model) - AIRS (Artificial Immune Recognition System)
Research perspective	- Accuracy - Efficiency	- Real-time - Versatility

4.1 사용 데이터에 따른 분류

이상거래 탐지의 경우 동일한 알고리즘을 사용하더라도 분석에 사용할 데이터에 따라 접근 방법이 다르므로 사용된 데이터에 따라 분류가 가능하다. 이상거래를 탐지하는 대상은 크게 네 분류로 신용카드 결제 데이터, 시뮬레이션 데이터, 은행 거래 데이터, 기타 데이터로 나눌 수 있다. Table 3.은 이상거래 탐지 연구들을 신용카드 결제 데이터, 시뮬레이션 데이터, 은행 거래 데이터, 기타 데이터로 분류한 것이다.

신용카드 결제 데이터는 카드사뿐만 아니라 PG사, 대형 매장 등에서 익명화된 데이터를 수집하기 때문에 이상거래 탐지 연구에 가장 많은 응용이 이루어진다. 신용카드 결제 데이터는 크게 결제 매장 데이터, 카드 정보, 결제 정보, 이용자 정보 등으로 구성된다. Whitrow 등은 2005년에 발생된 신용카드 거래데이터를 이용하여 이상거래 탐지 시 거래 집계(aggregation) 기간을 1일, 3일, 7일로 나누어 결과를 비교하였다[11]. Sanchez 등은 신용카드사기 거래를 탐지하기 위해 2002년부터 2003년 사이에 칠레 백화점에서 발생된 신용카드 거래 데이터를 활용하였다. 이용자의 개인정보만을 이용한 방식과 이용자의 개인정보 및 거래 데이터를 모두 사용하는 방식으로 나누어 진행하였으며 사용자의 성별, 나이, 거래 지역, 카드 발급 시기, 거래 금액, 거래 빈도 등이 사기 거래 사용자와 관련성이 높음을 확인하였다[12]. Bhattacharyya 등은 2006년부터 2007년 사이에 발생된 국제 신용카드 거래 데이터를 일정

Table 3. Classification based on data set

Data set	References	Description	Representative research
Credit card	[7-23]	- Many studies use the credit card transaction dataset. - Industries, avg. transaction value, and etc. are utilized by finding trading patterns	- “Transaction Aggregation as a Strategy for Credit Card Fraud Detection”[11]
Banking transaction	[24-26],[33]	- It consists of not only billing data but also financial status, credit card information, and the user information.	- “Fraudulent Electronic Transaction Detection using Dynamic KDA Model”[24]
Simulation data set	[27-33]	- Compared the efficiency between each other data mining algorithms and suggest optimal detection algorithm - The proposed methods should be to evaluate the performance via actual data.	- “Credit Card Fraud Detection Using Hidden Markov Model”[28]
etc.	[34-38]	- Mobile phone call records - In-game data - Peer-to-peer lending data	- “Automatic Detection of Compromised Accounts in MMORPGs”[38]

Table 4. Definition of variables about credit card transaction data[23]

Variable	Type	Definition
POS_no	char	POS terminal number
account_no	char	Account number
trans_date	date	Transaction date
trans_time	time	Transaction time
Trans_amt	float	The amount of a credit card transactions.
MCC	int	Merchant Category code
card_type	int	Card type
exp_date	date	Expiration Date
C1	float	The total amount of transactions of this card in the same day
C2	int	The number of transactions of this card in the same day
C3	int	The failure number of transactions of this card in the same day
C4	int	The average transaction amount of this card in the same day
C5	float	The total amount of transactions of this card within five days
C6	int	The number of transactions of this card within five days.
C7	int	The failure number of transactions of this card within five days.
C8	float	The average amount of trans. of this card within five days.

기간동안 집계하여 추가 필드를 생성한 후 데이터 마이닝 기법을 사용하여 탐지 능력을 평가하였다. 학습에 사용할 거래 데이터 셋에 정상 거래와 이상 거래의 불균형(unbalanced)을 샘플링(sampling) 기법으로 해결하였다[13]. Shen 등은 2005년부터 2006년 사이에 발생된 신용카드 데이터를 이상거래 탐지 연구에 사용하였으며 계좌 정보, 거래 일시, 카드 형태, 만료 일자, 상인 분류 코드 등 데이터 마이닝에 사용된 피쳐를 함께 정리하였다[23]. 해당 연구에서 사용된 신용카드 거래 데이터 셋을 구성하는 변수들에 대한 자세한 설명을 Table 4.에 정리하였다.

Table 5. Definition of variables about internet banking transaction data[26]

Variable	Type
Customer_ID	char
Transaction_date	date
Transaction_time	time
Transfer	boolean
ARS_Authentication	char
Transfer_Amount	int
Using_device	char
OS_ID	char
Access_IP	char
Country_code	char
VPN_Country	char
MAC	char

은행 거래 데이터는 입금 및 이체 데이터, ATM 기기를 통한 출금 데이터, 온라인 뱅킹(online banking) 데이터, 주식 및 증권 데이터 등으로 구성된다. 은행 거래 데이터를 통한 이상거래 탐지 시 결제 데이터를 재정 상태, 신용 정보 등 이용자의 정보와 연계하여 분석이 가능하다. Vadoodparast 등은 약 360만 건의 이체 데이터 및 은행에서 사용하는 Rapidminer 모델링 도구를 이용해 이상 거래를 탐지하는 방법론을 제안하였다. 논문에서 제안한 KDA 모델은 온라인 동적 모델링 및 오프라인 모드에서 80% 이상의 사기 거래를 탐지하였다[24]. Liu 등은 CSMAR(China Stock Market & Accounting Research) 데이터 셋을 여러 데이터 마이닝 알고리즘에 적용하고, 결과를 통해 최적의 이상거래 탐지 모델을 제시하였다[25]. 우리의 기존 연구에서는 국내 모 은행의 전자금융 사고 데이터를 바탕으로 고객의 패턴 정보와 프로파일 정보를 도출하고 이를 통해 탐지 룰을 설정하였다[26]. 해당 연구에서 사용된 인터넷 뱅킹 데이터 셋을 구성하는 변수들의 정의를 Table 5.에 정리하였다.

시뮬레이션 데이터는 실제로 사용된 결제 데이터를 구할 수 없는 경우 실제 환경과 유사한 시나리오를 구성하여 생성할 수 있다. 범용적인 이상거래 탐지 알고리즘을 연구하거나 방법론을 제안하기 위해 주로 쓰이지만, 정상거래와 이상거래 사이에 인위적인 차이가 존재할 수 있다. 따라서 실제 거래 데이터

에 제안된 방법론을 적용하기 전에는 실제 성능을 평가하기 어려울 수 있다. Nicholas Patterson 등은 가상재화 도용탐지 알고리즘을 사용하는 탐지 프레임워크(framework) 설계를 제안하였다. 탐지 알고리즘을 계정침입 탐지와 가상재화 부정거래 탐지 두 단계로 나누어 6개의 시나리오 데이터에 적용하였다[29]. K. R. Seeja 등은 익명화 처리된 불균형 거래 데이터 셋에서 사기 또는 정상 거래 패턴인지 판별하는 매칭 알고리즘(matching algorithm)을 제안하였다[32].

그 외에 데이터 마이닝 기법을 게임 내 거래 데이터 및 통화기록 데이터 등에 적용할 수 있다. Hee Yeon 등은 모바일 뱅킹에서의 이상 거래 탐지를 위해 거래 데이터가 아닌 스마트폰의 터치스크린 입력 데이터를 데이터 마이닝 알고리즘에 적용하였다. 터치 이벤트 및 스크롤 이벤트에 대해 입력 시작 좌표, 끝 좌표, 손가락 기울기, 스크롤 속도 등을 측정하여 데이터 마이닝 알고리즘에 적용하고 성능이 우수함을 보였다[34]. 크라우드 펀딩(crowdfunding) 혹은 P2P(Peer-to-Peer) 대출은 소셜 네트워크의 발달로 금융 기관의 개입 없이 이용자들 사이에 돈을 빌려주거나 빌리는 시스템이다. Milad Malekipirbazari 등은 2012년부터 2014년 사이에 LendingClub에서 발생한 거래 데이터를 kNN, Logistic Regression, SVM, Random Forest 알고리즘에 적용하여 이상거래를 탐지하였다[35]. Jehwan Oh 등은 실제 온라인 게임 내의 퀘스트, 로그인, 거래 데이터를 통계 및 데이터 마이닝 기법에 적용하였다. 제시한 방법론은 p-value를 통해 유저들의 과거 행동들과 최근 행동의 변화를 확인하고 탈취된 계정을 탐지하였다[36].

4.2 데이터 마이닝 알고리즘에 따른 분류

이상거래 탐지 연구는 각 연구에서 제안한 방법에 따라 다양한 알고리즘이 사용 가능하다. 본 절에서는 이상거래 탐지 연구를 사용된 데이터 마이닝 알고리즘에 따라 분류하였다. Table 6.은 이상거래 탐지 연구들을 Decision Tree, Random Forest, Logistic Regression, SVM, ANN, SOM, Naive Bayes, Bayesian Network, Association Rule, kNN, Discriminant Analysis, HMM, AIRS 알고리즘으로 분류한 것이다.

Decision Tree 알고리즘은 의사결정 규칙을 트리 구조로 도표화하여 가장 좋은 예측인자에 해당하는 속성을 루트 노드에 배치하고 하위 가지로 분할해 하나의 클래스에 대응하게 하는 과정을 반복하여 진행한다. Aihua Shen 등은 신용카드 거래 데이터에 Decision Tree, Neural Network, Logistic Regression 알고리즘을 적용하여 성능을 비교하고, 이상거래 탐지에 가장 좋은 모델을 선택할 수 있는 프레임워크를 제안하였다[23].

Random Forest는 의사결정 트리를 기본 분류기로 사용하는 앙상블 학습 알고리즘 중 하나이다. 훈련 데이터를 통해 기본 분류기들의 집합을 구성하고 각 분류기의 결과에 대한 투표를 통해 분류를 수행한다. Chengwei Liu 등은 신용카드 거래 데이터 중 7가지의 피쳐를 선택해 회귀 모델을 구축하고 Decision Tree, SVM, kNN, Logistic Regression, Random Forest 알고리즘에 적용하여 Random Forest의 성능이 우수함을 보였다[25].

Logistic Regression은 두 집단 이상으로 나누어진 데이터를 개별 관측치를 통해 분류하거나 예측하는 알고리즘이다. 종속 변수가 두 가지일 때 특히 성능이 좋아 이상거래 탐지 연구에 자주 사용된다. ANN은 인간 두뇌의 신경세포를 모방하여 마디(node)와 고리(link)로 인공 신경망을 구성하고 반복 학습과정을 거쳐 데이터에 내재되어 있는 패턴을 찾아내는 모델링 알고리즘이다. Ganesh Kumar 등은 신용카드 거래 데이터를 Logistic Regression과 ANN에 적용하여 이상거래를 탐지하였으며 분류 과정을 상세하게 기술하였다[22].

SVM은 커널 함수를 사용하여 비선형 문제를 선형 문제로 전환하여 판별을 수행하며 학습에 오랜 시간이 걸리는 대신 정확도가 높아 많은 이상거래 탐지 연구에 사용되었다. Sharmila Subudhi 등은 휴대전화의 통화 기록에서 이상 행위를 탐지하기 위한 방법론을 제안하고 SVM, OC-SVM(One Class-SVM), QS-SVM(Quarter Sphere-SVM) 등에 적용하여 정확도, 분석 시간으로 성능을 평가하였다[38].

SOM은 주어진 입력 패턴을 자기 스스로 학습하여 입력 벡터에 존재하는 다양한 속성간의 관계를 계산하여 숨겨진 패턴을 찾는 알고리즘이다. Jon T.S. Quah 등은 약 2억 명 이상의 고객을 보유한 싱가폴의 은행 데이터를 SOM 알고리즘에 적용하여

Table 6. Classification based on data mining algorithm

Algorithm	References	Application class	Representative research
Decision Tree	[9],[20],[23], [25-26]	Classification, Regression	- “Application of Classification Models on Credit Card Fraud Detection”[23]
Random Forest	[11],[13],[19],[25],[31-32], [35]	Classification, Clustering, Regression	- “Financial Fraud Detection Model: Based on Random Forest”[25]
Logistic Regression	[11],[13], [22-23],[25], [32],[35]	Classification, Regression, Prediction,	- “Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit”[22]
SVM(Support Vector Machine)	[11],[13],[19],[25],[32], [34-35],[38]	Classification, Regression	- “Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks ”[38]
ANN(Artificial Neural Network)	[8-9], [13-15], [19-20], [22-23]	Classification, Regression, Prediction	- “Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit”[22]
SOM(Self-Organization Map)	[10],[27]	Clustering	- “Real-time Credit Card Fraud Detection using Computational Intelligence”[10]
Naive Bayes	[9],[11],[20], [31-32]	Classification, Clustering	- “Credit Card Fraud Detection with Artificial Immune System”[9]
Bayesian Network	[8-9],[14]	Classification, Prediction	- “Credit Card Fraud Detection using Bayesian & Neural Network”[8]
Association Rule	[12]	Outlier detection	- “Association Rules applied to Credit Card Fraud Detection”[12]
kNN(k-Nearest Neighbor)	[11],[17],[24],[31-32], [34-35]	Clustering, Regression	- “Conditional Weighted Transaction Aggregation for Credit Card Fraud Detection”[31]
Discriminant Analysis	[11]	Classification, Clustering	- “Transaction Aggregation as a Strategy for Credit Card Fraud Detection”[11]
HMM(Hidden Markov Model)	[14],[16],[28],[30],[37]	Classification	- “Study of Hidden Markov Model in Credit Card Fraudulent Detection”[16]
AIRS(Artificial Immune Recognition System)	[9],[18]	Classification, Prediction	- “Credit Card Fraud Detection with Artificial Immune System”[9]

실시간 이상거래 탐지 연구를 수행하였다. 적절한 뉴런의 수와 반복 학습을 통해 이상거래 탐지 시스템을 계층화하였고 새로운 사기 경향에 적용 가능함을 보였다[10]. Vladimir Zaslavsky 등은 SOM 알고리즘에 대해 설명하고 알고리즘을 구성하는 변수들을 이상거래 탐지에 활용하기 위한 방법론을 제시하였다. 한 달에 100건 이하의 거래를 수행하는 결제를 수행하는 소비자로부터 결제 패턴을 분석하고 이로부터 이상 행위 탐지가 가능함을 보였다[27].

Naive Bayes는 Bayes 이론을 이용한 확률 기반의 분류 방법으로 전체적인 분류 정확도를 최대화하거나 특정 클래스에 속하는 레코드를 식별하는데 사용된다. Manoel Fernando Alonso Gadi 등은 신용카드 사기거래를 높은 효율로 탐지하기 위한 연구를 수행하였다. Naive Bayes, Neural Network, Bayesian Network, Decision Tree, AIRS 알고리즘을 사용하였으며 AIRS에 최적화된 파라미터 설정을 통해 저비용으로 높은 탐지

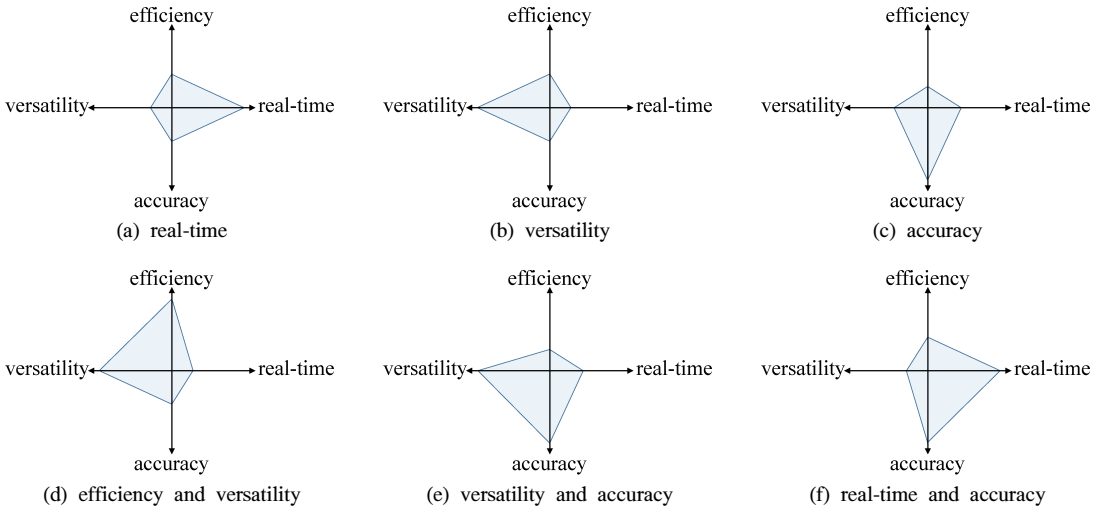

Fig. 2. Description of the research perspective

가 가능함을 보였다[9].

kNN 알고리즘은 가장 간단한 비모수 방식의 기계 학습 알고리즘 중 하나로 빠른 시간 안에 데이터의 거리를 측정하여 각 데이터를 k개의 집합 중 하나로 분류한다. Venkata Ratnam 등은 신용카드 스트림 데이터에 kNN을 응용한 SODRNN 알고리즘을 프로그램으로 구현하고 이상거래 분류 시 제안한 방법이 높은 성능과 메모리 효율을 가짐을 보였다[17].

4.3 연구 관점에 따른 분류

본 절에서는 이상거래 탐지 연구 관점을 실시간 탐지 가능성, 범용성, 정확도, 분류 효율로 분류한다. Fig. 2.는 네 가지 연구 관점을 그림으로 나타낸 것이다. 시뮬레이션 데이터를 이용한 연구는 주로 실제 거래 데이터에 적용할 수 있게 범용성에 초점을 두고 있다. 반면 실시간으로 학습하고 분류에 적용 가능한 알고리즘을 사용한 경우, 간단하며 효율이 좋은 알고리즘을 사용한 경우 실시간 탐지를 고려할 수 있다. 그러나 실시간 탐지에 사용할 피쳐는 시스템에 따라 충분한 가공 과정과 알고리즘 최적화 작업이 필요하므로 범용성과는 거리가 멀다. 또한 거래 데이터를 면밀히 분석하여 이상거래 분류를 높은 정확도로 수행하려는 경우 분류에 오랜 시간이 소요되므로 정확도와 분류 속도는 상반된 관점이다.

Table 7.은 연구 관점에 따라 이상거래 탐지 연구를 분류한 것이다. (a)는 실시간 이상거래 탐지에 관한 연구를 정리한 것이다. 실시간 탐지의 경우 순간적으로 정확한 판단을 내려야 하고 데이터 마이닝 알고리즘의 특성을 고려해야 한다. T.S. Quah 등은 SOM 알고리즘을 이용해 초기 인증 계층, 검사 계층, 위험 점수 평가와 행위 분석을 하는 핵심 계층, 추가 검토 계층 등으로 탐지 메커니즘을 구분하여 실시간 탐지를 구현하였다[10].

(b)는 범용성 관점에서 데이터 마이닝 기법을 이상거래 탐지에 적용할 수 있음을 보인 연구를 정리한 것이다. 이에 속한 연구들은 주로 거래 데이터를 데이터 마이닝 알고리즘에 적용하기 위한 전처리 과정을 상세히 기술하고 성능을 확인한다. Andrea Dal Pozzolo 등은 신용카드사기 탐지 과정에서 발생하는 데이터 불균형, 비정지 분포(non-stationarity) 평가에 대한 문제점과 해결책을 제시하였다. 비정지 분포는 카드 소유자와 사기꾼의 지출 행위(spending behavior)가 진화하기 때문에 결제 패턴이 수시로 변하는 문제를 말하며, static approach, updating approach, forgetting genuine approach와 같은 방법론을 제안하였다.[19]

(c)는 정확도 관점에서 전처리된 결제 데이터를 여러 데이터 마이닝 알고리즘에 적용하여 이상거래 분류를 수행하고 결과 분석을 통해 이상거래 탐지율을 향상시킨 연구를 정리한 것이다. Abhinav Srivastava 등은 신용카드 거래 금액을 시간 순서

Table 7. Classification based on research perspective

Class	Research point	References	Representative research
(a)	real-time	[10]	- “Real-time Credit Card Fraud Detection using Computational Intelligence”[10]
(b)	versatility	[15],[19],[21],[26-27], [29],[33],[34],[36]	- “Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective”[19]
(c)	accuracy	[7-9],[12],[18],[20], [22],[23], [25],[28],[35],[37]	- “Credit Card Fraud Detection using Hidden Markov Model”[28]
(d)	efficiency and versatility	[14],[17],[30]	- “Implement Credit Card Fraudulent Detection System using Observation Probabilistic in Hidden Markov Model”[30]
(e)	versatility and accuracy	[11],[13],[24],[31],[38]	- “Data Mining for Credit Card Fraud: A Comparative Study”[13]
(f)	real-time and accuracy	[16]	- “Study of Hidden Markov Model in Credit Card Fraudulent Detection”[16]

에 따라 시퀀스로 작성하고 HMM 알고리즘을 적용하여 이상거래를 탐지하는 방법론을 제안하였다. 시뮬레이션 데이터를 이용해 실험을 진행하고 성능 평가를 통해 사기 거래 탐지에 가장 적합한 HMM 파라미터를 확인하였다[28].

(d)~(f)는 기존에 진행된 이상거래 탐지 연구를 응용하여 더 나은 성능의 이상거래 탐지 알고리즘을 제안한 연구를 정리한 것이다. 탐지 효율은 이상거래를 탐지하기 위한 알고리즘 수행 시간을 짧게 개선하거나 분류에 드는 비용, 오탐(false-positive) 및 미탐(false-negative)을 해결하는데 드는 비용을 줄이는 것으로 달성할 수 있다. 그러나 탐지 효율만을 고려한 이상거래 탐지 연구는 진행되지 않았다. 탐지 효율만을 고려하고 정확도 향상을 포기하는 것은 이상거래 탐지 시스템의 목적과는 거리가 멀기 때문인 것으로 보인다. 또한 실시간 탐지는 실제 데이터를 통한 분석과 함께 실제 거래 시스템을 통한 검증을 거쳐야 한다는 점 때문에 많은 연구가 진행되지는 않았다.

V. 고 찰

Fig. 3.은 본 논문에서 조사한 이상거래 탐지 연구를 분류 방법에 따라 연도별 이상거래 탐지 연구 분포를 나타낸 것이다. (a)는 데이터 마이닝 알고리즘에 사용된 결제로그 데이터 셋에 따라 연구별 이상거래 탐지 연구 분포를 나타낸 것이다. 신용카드 데이터를 이용한 거래가 가장 많이 발생하였다. 신용카드 결제 데이터는 카드사, PG사, 온라인 및 오프라인 매장 등 다양한 출처로부터 얻을 수 있기 때문에 이를 이용한 연구가 용이하고, 결제 레코드를 구성하는 피쳐가 유사하므로 다른 시스템에 적용하기 용이한 장점이 있기 때문으로 볼 수 있다.

(b)는 13가지의 데이터 마이닝 알고리즘 중 가장 많이 사용된 상위 5개 알고리즘에 따라 연도별 이상거래 탐지 연구 분포를 나타낸 것이다. Random Forest는 분류, 클러스터링, 회귀 모두 효과적일 뿐만 아니라 특정 학습데이터에 민감한 문제를 앙상블 기법으로 극복 가능한 장점 때문에 이상거래 탐지 연구에서 많이 사용된 것으로 보인다. Logistic Regression은 종속 변수와 한 개 이상의 독립 변수 사이의 관계 분석을 통해 기존의 데이터를 설명하고 새로운 데이터를 예측하는 장점이 있어 연구에 많이 사용된 것으로 보인다. SVM은 binary 분류에 효과적이고 과적합 위험을 최소화해주는 장점이 있다. 커널 함수 때문에 분류 시간이 오래 걸리는 단점이 있었지만 하드웨어 성능의 발달로 인해 현재 가장 널리 사용되는 기계학습 알고리즘 중 하나이다. 신용카드 결제 데이터에 Random Forest, Logistic Regression, SVM 등 세 알고리즘을 동시에 비교한 연구가 진행되었다[13].

(c)는 연구 관점에 따른 연도별 연구 분포를 나타낸 것이다. 정상 이용자의 불편함을 줄이고 새로운 사기 거래 패턴을 탐지하기 위해 이상 거래를 탐지 시 정확도에 중점을 두는 연구가 가장 많이 진행되었다. 이러한 연구들은 주로 동일한 데이터 셋에 여러

데이터 마이닝 알고리즘을 적용하여 결과를 통해 성능을 비교하였다. 또한 매 해를 거듭할수록 시스템의 성능이 비약적으로 향상되는 점 또한 정확도 관점에서 연구가 진행되는 것에 영향을 미친 것으로 보인다.

VI. 결 론

전자금융거래 기술의 발달과 대중화로 인해 이를 이용한 사기행위가 점차 발전하고 있다. 신용카드 및 은행 거래에서 발생된 이상거래는 금전적 손실과 직접적으로 연결되므로 이를 탐지하기 위한 대응 방안이 필요하다. 많은 금융 기관이 이용자의 재산을 보호하기 위해 이상거래 탐지 시스템을 구축하고 있다. 하지만 이상거래 패턴이 다양하고 새로운 사기 기법들이 등장하기 때문에 여러 관점에서 탐지 연구가 이루어져야 한다.

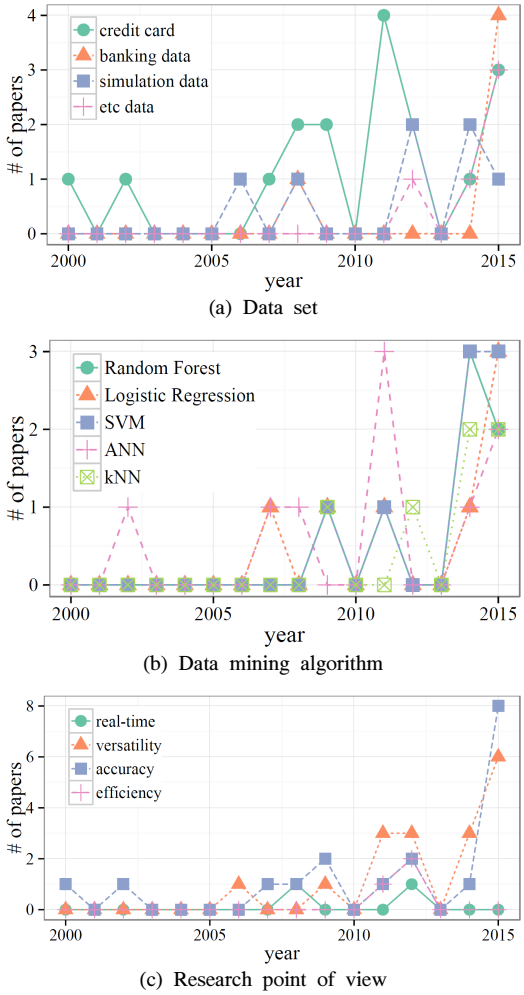

Fig. 3. Distribution of articles by research year

본 논문에서는 결제로그 분석 및 데이터 마이닝을 이용한 이상거래 탐지 연구들을 조사하였으며, 데이터 마이닝 기법에 대해 정리하고 관련 연구들을 사용된 데이터 셋, 데이터 마이닝 알고리즘, 연구 관점 등 3가지 분류기준으로 분류하였다. 또한 분류기준별로 연도별 연구 분포를 통해 최신 연구 경향을 제시하였다. 우리는 본 연구를 통해 이상거래 탐지를 시도하려는 다양한 관점의 연구들을 조금 더 쉽게 이해할 수 있으며, 제시된 데이터 마이닝 알고리즘 적용 방안을 참고하여 더 나은 성능을 제공하는 FDS를 구축할 수 있으리라 기대한다. 비트코인과 같은 신종 화폐의 등장과 크라우드펀딩 활성화 등 새로운 결제 환경이 조성되고 있기 때문에 기존에 진행된 연구를 바탕으로 향후 다양한 환경에서 대응 방안이 마련되어야 할 것이다.

References

[1] Gyusoo Kim and Seulgi Lee, "2014 Payment Research," Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[2] Yunhwa Lim, "Research of Statistics and Countermeasure about Credit Card Fraud," The Credit Finance Research Institute, Vol. 2015, No. 16, Nov. 2015.

[3] Financial supervisory service, "Electronic Fraud Prevention Services," May. 2013.

[4] E.W.T. Nagi, Yong Hu, H.Y. Wong, Yijun Chen, Xin Sun, "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, Feb. 2011.

[5] Jha, Sanjeev, J. Christopher Westland, "A Descriptive Study of Credit Card Fraud Pattern," Global Business Review, Vol. 14, No. 3, pp. 373-384, 2015.

[6] Edge, Michael Edward, Pedro R. Falcone Sampaio, "A Survey of Signature based Methods for Financial Fraud Detection,"

Computers & Security, Vol. 28 No. 6, pp. 381-394. 2009.

〔7〕 Peter J. Bentley, Jungwon Kim, Gil-Ho Jung, Jong-Uk Choi, "Fuzzy Darwinian Detection of Credit Card Fraud," The 14th Annual Fall Symposium of the Korean Information Processing Society, Oct. 2000.

〔8〕 Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard manderick, "Credit Card Fraud Detection Using Bayesian and Neural Networks," Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies, Jan. 2002.

〔9〕 Manoel Fernando Alonso Gadi, Xidi Wang, Alair Pereira do Lago, "Credit Card Fraud Detection with Artificial Immune System," Artificial Immune Systems, pp. 119-131, 2008.

〔10〕 Jon T.S. Quah, M. Sriganesh, "Real-time Credit Card Fraud Detection using Computational Intelligence," Expert Systems with Applications, Vol. 35, No. 4, pp. 1721-1732, Nov. 2008.

〔11〕 C. Whitrow, D.J. Hand, P. Juszczak, D. Weston, N.M. Adams, "Transaction Aggregation as a Strategy for Credit Card Fraud Detection," Data Mining and Knowledge Discovery, Vol. 18, No. 1, pp. 30-55, Feb. 2009.

〔12〕 D. Sáncheza, M.A. Vilaa, L. Cerdaa, J.M. Serranob, "Association Rules applied to Credit Card Fraud Detection," Expert Systems with Applications, Vol. 36, No. 2, pp. 3630-3640, Mar. 2009.

〔13〕 Siddhartha Bhattacharyyaa, Sanjeev Jhab, Kurian Tharakunnelc, J. Christopher, "Data Mining for Credit Card Fraud: A Comparative Study," Decision Support System, Vol. 50, No. 3, pp. 602-613, Feb. 2011.

〔14〕 S.B.E. Raj, A.A. Portia, "Analysis on Credit Card Fraud Detection Methods," Computer, Communication and Electrical Technology (ICCCET) 2011 International Conference on, pp. 152-156, Mar. 2011.

〔15〕 Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection using Neural Network," International Journal of Soft Computing and Engineering (IJSCE), pp. 33-38, 2011.

〔16〕 V. Bhusari, S. Patill, "Study of Hidden Markov Model in Credit Card Fraudulent Detection," International Journal of Computer Applications, Vol. 20, No. 5, Apr. 2011.

〔17〕 V.R. Ganji, S.N.P. Mannem, "Credit Card Fraud Detection using Anti-k Nearest Neighbour Algorithm," International Journal on Computer Science and Engineering, Vol. 4, No. 6, pp. 1035-1039, Jan. 2012.

〔18〕 Neda Soltani, Mohammad Kazem Akbari, Mortaza Sargolzaei Javan, "A New User-Based Model for Credit Card Fraud Detection Based on Artificial Immune System," Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on IEEE, May. 2012

〔19〕 A.D. Pozzoloa, O. Caelenb, Y.A.L. Borgnea, S. Waterschootb, G. Bontempia, "Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective," Expert Systems with Applications, Vol. 41, No. 10, pp. 4915-4928, Aug. 2014.

〔20〕 Nader Mahmoudi, Ekrem Duman, "Detecting Credit Card Fraud by Modified Fisher Discriminant Analysis," Expert Systems with Applications, Vol. 42, No. 5, pp. 2510-2516, Apr. 2015.

〔21〕 Ivo Correia, Fabiana Fournier, Inna Skarbovsky, "The Uncertain Case of Credit Card Fraud Detection," Proceedings of the 9th ACM International

Conference on Distributed Event-Based Systems, pp. 181-192, 2015.

[22] Ganesh Kumar.Nune and P.Vasanth Sena, "Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit," International Journal of Computer Science and Network Security, Vol. 15, No. 9, Sep. 2015.

[23] Aihua Shen, Rencheng Tong, Yaochen Deng, "Application of Classification Models on Credit Card Fraud Detection," Service Systems and Service Management of the 2007 IEEE International Conference, pp. 1-4, Jun. 2007.

[24] Massoud Vadoodparast, Abdul Razak Hamdan, Hafiz, "Fraudulent Electronic Transaction Detection using Dynamic KDA Model," International Journal of Computer Science and Information Security, Vol. 13, No. 3, pp. 90-99, Mar. 2015.

[25] Chengwei Liu, Yixiang Chan, Syed Hasnain Alam Kazmi, Hao Fu, "Financial Fraud Detection Model: Based on Random Forest," International Journal of Economics and Finance, Vol. 7, No. 7, pp. 178-188, 2015.

[26] Jae Hoon Park, Huy Kang Kim, Eunjin Kim, "Effective Normalization Method for Fraud Detection Using a Decision Tree," Journal of the Korea Institute of Information Security and Cryptology, Vol. 25, No. 1, pp. 133-146, Feb. 2015.

[27] Vladimir Zaslavsky and Anna Strizhak, Credit card fraud detection using self-organizing maps, Information and Security, Vol. 18, pp. 48-63, 2006.

[28] Abhinav Srivastava, Amlan Kundu, Shamik Sural, "Credit Card Fraud Detection using Hidden Markov Model," Dependable and Secure Computing, Vol. 5, No. 1, pp. 37-48, Jan. 2008.

[29] Nicholas Patterson, Michael Hobbs, Jemal Abawajy, "Virtual Property Theft Detection Framework: An Algorithm to Detect Virtual Property Theft in Virtual World Environments," Trust, Security and Privacy in Computing and Communications on IEEE, pp. 177-184, Jun. 2012.

[30] Ashphak Khan, Tejpal Singh, Amit Sinhal, "Implement Credit Card Fraudulent Detection System using Observation Probabilistic in Hidden Markov Model," Engineering (NUiCONE), 2012 Nirma University International Conference on IEEE, pp. 1-6, Dec. 2012.

[31] Wee-Yong Lim, Amit Sachan, Vrizlynn Thing, "Conditional Weighted Transaction Aggregation for Credit Card Fraud Detection," Advances in Digital Forensics X. pp. 3-16, 2014.

[32] K.R. Seeja and Masoumeh Zareapoor, "FraudMiner: A Novel Credit Card Fraud Detection Model based on Frequent Itemset Mining," The Scientific World Journal, Vol. 2014, Sep. 2014.

[33] Luigi Coppolino, Salvatore D'Antonio, Valerio Formicola, Carmine Massei, Luigi Romano, "Use of the Dempster-Shafer Theory for Fraud Detection: The Mobile Money Transfer Case Study," Intelligent Distributed Computing VIII, pp. 465-474, 2015.

[34] Hee Yeon Min, Jin Hyung Park, Dong Hoon Lee, In Seok Kim, "Outlier Detection Method for Mobile Banking with User Input Pattern and E-finance Transaction Pattern," Journal of Korean Society for Internet Information, Vol. 15, No. 1, pp. 157-170, Feb. 2014.

[35] Milad Malekipirbazari and Vural Aksakalli, "Risk Assessment in Social Lending via Random Forests," Expert Systems with Applications, Vol. 42, No.

10, pp. 4621-4631, Jun. 2015.
〔36〕 Jehwan Oh, Zoheb Hassan Borbora, Jaideep Srivastava, "Automatic Detection of Compromised Accounts in MMORPGs," Social Informatics (SocialInformatics), International Conference on. IEEE, pp. 222-227, Dec. 2012.
〔37〕 Constantinos S. Hilas1, Paris A. Mastorocostas1, Ioannis T. Rekanos, "Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A Case Study," Vol. 9, No. 4, pp. 1709-1718, Jan. 2015.
〔38〕 Sharmila Subudhia, Suvasini Panigrahib, "Quarter-Sphere Support Vector Machine for Fraud Detection in Mobile Telecommunication Networks," Procedia Computer Science, Vol. 48, pp. 353-359, 2015.

〈저자소개〉

정 성 훈 (Seong Hoon Jeong) 학생회원
2015년 2월: 충북대학교 정보통신공학부 졸업
2015년 3월~현재: 고려대학교 정보보호학과 석사과정
〈관심분야〉 네트워크 보안, 시스템 보안, 데이터 마이닝

김 하 나 (Hana Kim) 학생회원
2013년 2월: 서울여자대학교 정보보호학과 졸업
2015년 8월: 고려대학교 정보보호학과 석사
2015년 9월~현재: 고려대학교 정보보호학과 박사과정
〈관심분야〉 온라인게임 보안, 데이터 마이닝

신 영 상 (Youngsang Shin) 정회원
1998년 2월: 부산대학교 컴퓨터공학 학사
2000년 2월: 부산대학교 컴퓨터공학 석사
2004년 5월: University of Wisconsin - Madison, Computer Science 석사
2011년 8월: Indiana University - Bloomington, Computer Science 박사
2011년 12월~현재 한국인터넷진흥원 책임연구원
〈관심분야〉 네트워크/시스템 침입방지, 클라우드 보안, 모바일 보안, 웹 보안, 금융 보안

이 태 진 (Tae Jin Lee) 정회원
2003년 : POSTECH 컴퓨터공학 학사 졸업
2008년 : 연세대학교 컴퓨터공학 석사 졸업
2003년 1월~현재 : 한국인터넷진흥원 팀장
〈관심분야〉 : 악성코드, 네트워크 보안, 시스템 보안

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업공학과 석사
2009년 2월: KAIST 산업및시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~2014년 12월: 고려대학교 정보보호대학원 조교수
2015년 1월~현재: 고려대학교 정보보호대학원 부교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌식