JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 26 No. 4, pp. 93-103, April 2021
https://doi.org/10.9708/jksci.2021.26.04.093

A Classification Model for Illegal Debt Collection Using Rule and Machine Learning Based Methods

Tae-Ho Kim*, Jong-In Lim*

*Student, Graduate School of Information Security, Korea University, Seoul, Korea
*Professor, Graduate School of Information Security, Korea University, Seoul, Korea

[Abstract]

Despite the efforts of financial authorities in conducting the direct management and supervision of collection agents and bond-collecting guideline, the illegal and unfair collection of debts still exist. To effectively prevent such illegal and unfair debt collection activities, we need a method for strengthening the monitoring of illegal collection activities even with little manpower using technologies such as unstructured data machine learning. In this study, we propose a classification model for illegal debt collection that combine machine learning such as Support Vector Machine (SVM) with a rule-based technique that obtains the collection transcript of loan companies and converts them into text data to identify illegal activities. Moreover, the study also compares how accurate identification was made in accordance with the machine learning algorithm. The study shows that a case of using the combination of the rule-based illegal rules and machine learning for classification has higher accuracy than the classification model of the previous study that applied only machine learning. This study is the first attempt to classify illegalities by combining rule-based illegal detection rules with machine learning. If further research will be conducted to improve the model's completeness, it will greatly contribute in preventing consumer damage from illegal debt collection activities.

▸ **Key words**: Illegal Debt Collection, Machine Learning, Classification Machine Learning, Text Mining, Document classification, Support Vector Machine

[요 약]

금융당국의 채권추심 가이드라인, 추심업자에 대한 직접적인 관리 감독 수행 등의 노력에도 불구하고 채무자에 대한 불법, 부당한 채권 추심은 지속되고 있다. 이러한 불법, 부당한 채권추심행위를 효과적으로 예방하기 위해서는 비정형데이터 기계학습 등 기술을 활용하여 적은 인력으로도 불법 추심행위에 대한 점검 등에 대한 모니터링을 강화 할 수 있는 방법이 필요하다. 본 연구에서는 대부업체의 추심 녹취 파일을 입수하여 이를 텍스트 데이터로 변환하고 위법, 위규 행위를 판별하는 규칙기반 검출과 SVM(Support Vector Machine) 등 기계학습을 결합한 불법채권추심 분류 모델을 제안하고 기계학습 알고리즘에 따라 얼마나 정확한 식별을 하였는지를 비교해 보았다. 본 연구는 규칙기반 불법 검출과 기계학습을 결합하여 분류에 활용할 경우 기존에 연구된 기계학습만을 적용한 분류모델 보다 정확도가 우수하다는 것을 보여 주었다. 본 연구는 규칙기반 불법검출과 기계학습을 결합하여 불법여부를 분류한 최초의 시도이며 후행연구를 진행하여 모델의 완성도를 높인다면 불법채권 추심행위에 대한 소비자 피해 예방에 크게 기여할 수 있을 것이다.

▸ **주제어**: 불법채권추심, 기계학습, 분류 기계학습, 텍스트 마이닝, 문서분류, 서포트 벡터 머신

• First Author: Tae-Ho Kim, Corresponding Author: Jong-In Lim
*Tae-Ho Kim (taeho@fss.or.kr), Graduate School of Information Security, Korea University
*Jong-In Lim (jilim@korea.ac.kr), Graduate School of Information Security, Korea University
• Received: 2021. 03. 03, Revised: 2021. 03. 23, Accepted: 2021. 04. 21.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

금융당국의 채권추심 가이드라인, 추심업자에 대한 직접적인 관리 감독 수행 등의 노력에도 불구하고 채무자에 대한 불법, 부당한 채권 추심은 지속되고 있는 상황이다. 특히, 금융회사 등은 연체 채권을 관리비용 절감, 재무건전성 개선, 유동성 확보 등을 위해 매각하고 있으나 가계신용대출 채권의 경우 대부분 대부업체가 매입하고 있어 대부업체의 채무자로 전락하고 과도한 추심에 노출되는 경우가 다수 발생하고 있는 실정이다. 정부가 법정 최고금리 인하와 정책서민금융을 확대하면서 대부업 전체 규모가 갈수록 줄어들고 있지만 대부업 수익성이 줄면서 제도권 대부업체가 대출 물량을 중단하거나 축소하고 있어 취약계층이 불법 사금융으로 내몰리는 '풍선효과'가 발생하고 있다는 우려가 있다. 대부업 시장은 규모가 줄고 있음에도 '20년 1~5월 기간 중 금융감독원에 불법 고금리, 채권추심 등 상담·신고 건수는 2,313건으로 전년 동기 대비 56.9% 증가한 것으로 나타나고 있다.[1]

Table 1. Results of the report of damage by the illegal private financial reporting center (daily average)

Classification	'19 Year	'20 Year				
		Jan	Feb	Mar	Apr	May
최고금리 위반	2.3	1.6	2.6	5.0	5.9	4.7
불법추심	1.6	2.4	2.9	2.9	3.1	2.0
미등록대부	10	16.2	12.4	12.0	16.9	15.4
불법대부광고	6	4.6	4.7	8.2	8.5	8.0
불법중개수수료	0.4	0.5	1.2	0.5	1.0	0.6

불법 채권추심으로 인한 폐해의 심각성은 사회 문제로 계속 대두되고 있다. 이러한 불법, 부당한 채권추심행위를 효과적으로 예방하기 위해서는 빅데이터, 기계학습 등 최신 기술을 활용하여 적은 인력으로도 불법 추심행위에 대한 점검 등 모니터링을 강화 할 수 있는 방법이 필요하다. 본 연구에서는 대부업체의 추심행위에 대한 녹취 파일을 입수하여 이를 텍스트 데이터로 변환하고 위법, 위규 행위를 분류하는 불법채권추심 분류 모델을 제안하였다. 채권추심 녹취 파일을 텍스트 데이터로 변환하고 이를 규칙기반 검출과 Support Vector Machine(SVM), Random Forest(RF), Convolutional Neural Network (CNN) 기계학습 알고리즘을 적용하여 불법채권 추심행위 여부를 분류하고 실제 각 알고리즘별로 얼마나 정확한 식별을 하였는지 성능분석을 하였다.

II. Preliminaries

1. Related works

빅데이터(Big Data)분석에 대한 수요와 관심은 2016년 알파고의 등장 이후 계속해서 높아지고 있다. 빅데이터 분석의 대상은 동영상, 이미지, 음성 등의 다양한 형태를 포함하며, 특히 정보전달의 대표적 수단인 텍스트를 분석하는 다양한 연구가 진행되고 있다. 그 중 텍스트 판별(Classification)은 텍스트 마이닝 응용 분야 중 하나로 다양한 산업에서 활발하게 사용되고 있는 대표적 기술이다. 텍스트 판별은 지정된 카테고리로 문서를 분류하는 기술로 문서는 그 특성에 따라 하나의 클래스 혹은 둘 이상의 클래스로 라벨을 붙일 수 있다. 하나의 문서에 하나의 종류가 할당되는 텍스트 분류문제를 "단일 레이블 분류"라고 하며, 문서에 둘 이상의 클래스가 할당된다면 이는 "다중 레이블 분류"다.[2] 텍스트 분류를 위한 다양한 방법과 절차가 있지만 크게 네 가지 과정을 따른다.[3] 첫 번째는 문서를 기계가 이해할 수 있는 형태로 변형하는 것으로 정해진 특성의 수로 문서의 표현 방법을 변형한다. Bag-Of-Words(BOW)는 흔히 쓰이는 단어기반의 문서표현 방법으로 이를 통해 수많은 단어로 이루어진 문서가 비교적 작은 단어 집합으로 표현된다. 두 번째는 특징선택(feature selection) 과정으로 Document Frequency, Information Gain 등 단어별 중요도를 점수 매겨 데이터 차원을 줄임으로써 문서 분류의 정확도와 분류 모델의 과적합을 막는다. 세 번째 과정은 특징변환(Feature Transformation)으로 특징선택과 같이 특성의 수를 줄이는 것에 그 목적이 있으며 Principal Component Analysis(PCA)는 가장 많이 쓰이는 방법론이다.[4] 고유값 분해를 통해 중요한 벡터공간에서 중요한 특성들을 선별한다. 특징선택과 특징변환을 완료했다면 Decision Tree(DT), Support Vector Machines (SVM) 등의 다양한 데이터 마이닝 방법론과 Convolutional Neural Network(CNN), Recurrent Neural Network(RNN) 등의 딥러닝 방법론을 활용하여 분류기(classifier)를 만든다.[5-7] 이러한 텍스트 판별기술은 다양한 분야에서 활용되고 있다. 소셜미디어 데이터의 증가로 2000년대 이후 문서로부터 감정을 분석하는 많은 연구가 진행됐다, 영화 리뷰의 긍·부정을 판별하고 판별결과와 영화 흥행간 상관관계 파악을 하거나 상품평에 나타난 감성표현 분류를 통해 마케팅에 활용하는 시도가 있었다.[8,9] 텍스트 판별 기술은 소비자의 니즈를 파악하는 것뿐만 아니라 소비자를 보호하기 위해서도 활용된다. 프로그램 개발 보안 취약점에

해당하는 패턴을 기계학습 시키고 분류하여 보안 약점 탐지에 활용하기도 하고 SNS에 다량으로 배포되는 불법금융광고를 판별하여 금융소비자를 보호한다.[10,11] 최근에는 CNN으로 실제 판례를 학습하여 문서 안의 내용이 적법한지를 판별하는 수준에 이르렀다.[12,13] STT(Speech To Text) 기술의 발전은 음성을 문자로 변형함으로써 텍스트 판별 기술의 활용 범위를 넓혔다. 금융상품을 판매하는 상담원과 고객 간 대화문을 STT 기술로 변환하여 CNN 모델로 학습하여 금융상품 불완전판매 여부를 판별하는 모델을 만들었다.[14] 본 연구는 음성데이터를 텍스트화하여 활용할 수 있는 분야로 불법채권추심을 기계학습과 규칙기반 기법을 결합한 분류 모델을 제안하였고 이러한 분류 모델이 기존에 연구된 기계학습만을 적용하는 기법 보다 분류 정확도가 더 높다는 것을 보여주었다.

2. Theoretical Background

1) Keyword Embedding

키워드는 텍스트 자료의 중요한 내용을 함축적으로 나타내는 단어나 문구를 의미한다. 키워드 임베딩은 기계학습을 위해 키워드를 실수 벡터로 변환하는 과정을 말한다. 보통 SVM, RF 학습에는 tf-idf를 사용하고, CNN 학습 시에는 Word2Vec를 사용한다.

- tf-idf

기계학습 과정에서 각각의 키워드는 문서 분류를 위한 특징(feature)으로 사용된다. 따라서 하나의 문서를 기계학습을 위한 벡터로 변환하기 위해서는 각 키워드를 벡터 내 약속된 위치에 대해 실수값으로 변환하는 과정이 필요하다. 실제 tf-idf는 키워드의 문서 내 빈도(tf : term frequency)와 키워드가 존재하는 문서의 빈도(df : document frequency)를 이용하여 계산되며, 아래 식은 다양한 계산 방법들 중 가장 간단한 형태를 나타내었다. (t: 키워드, d: 특정 문서, D: 전체 문서 집합)

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) =$$

$$t\text{의 문서 내 출현빈도} \times \log\left(\frac{\text{전체 문서 수}}{t\text{포함 문서 수}}\right)$$

- Word2Vec

Word2Vec는 NN(Neural Network)를 이용하거나 확률 분포를 이용하여 키워드의 연관관계를 기반으로 키워드를 실수 벡터로 임베딩하는 방법을 말한다. tf-idf와의 가장 큰 차이점은 tf-idf로 변환된 벡터는 모든 가능한 키워드의 개수가 벡터의 길이가 되지만, Word2Vec는 길이를 고정하고 학습하기 때문에 보다 작은 길이의 벡터로 변환 가능하다는 점이다. 그 밖에도 키워드의 연관관계를 이용하여 임베딩 함으로써 벡터 공간에 맵핑된 키워드들 간 의미 관계를 이용할 수 있다는 장점이 있으며, 임베딩 된 키워드가 의미를 내포함으로 인하여 딥러닝 학습 시에 좋은 성능을 보이게 된다.

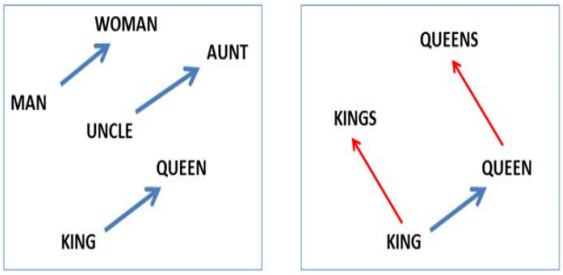

Fig. 1. Example of meaning relations among Word2Vec key words

[Fig 1]을 보면, 예시에 사용된 키워드는 모두 '남자'-'여자'의 관계를 가지고 있으며, 두 키워드들 간 방향 벡터는 서로 유사함을 확인할 수 있다. Word2Vec의 종류로는 Word2Vec 이외에도 Glove, Fast text 등이 있으며, 최근에도 연구가 활발히 진행되고 있다. 각 Word2Vec의 성능 평가는 의미 관계와 문법 관계 테스트 셋에 대한 정확도를 기반으로 한다. Word2Vec는 대부분 딥러닝의 입력 전처리 과정에 사용된다. 이는 딥러닝 학습 시 입력 특징(feature)의 개수가 학습하는 데이터의 수에 비해 너무 큰 값을 가지게 되면 제대로 된 학습이 이루어지기 어렵기 때문이다. SVM과 같은 기계학습 기법에는 보통 Word2Vec를 사용하지 않는데, 이는 일반적인 기계학습은 각 특징들을 서로 독립적인 입력값으로 가정하고 학습을 진행하기 때문이다. Word2Vec는 N 차원의 벡터를 M 차원으로 맵핑하기 때문에 임베딩 벡터 내 요소들끼리 서로 독립적인 학습이 불가능하다. 따라서 이후 성능분석 시에 SVM, RF 학습에는 tf-idf를 사용하고 CNN 학습 시에는 Word2Vec을 사용하였다.

2) Imbalance Dataset

분류기 학습에 사용되는 데이터는 각 문서별로 분류가 태깅되어 있는데, 이 때 각 분류별 문서의 개수가 차이가 많이 나는 경우를 imbalance dataset 이라 부른다. 이러한 불균형 데이터를 학습하게 되면, 다수의 문서와 맵핑되

는 분류에 대해 overfitting 되어 학습되기 때문에, 실제 적용 시 해당 분류로 더 많이 예측하도록 분류기가 학습되어진다. 하지만, 본 연구의 테스트 데이터 셋은 학습 및 테스트 데이터 셋의 비율이 동일하므로 분류기 학습 시 특정 분류로 편향되어 학습되는 것을 막아야 한다. 대표적인 방법으로 bootstrap, oversampling 등이 있는데, bootstrap을 사용하기에는 불법 분류의 개수가 너무 작아 oversampling을 적용하였다. [Fig 2]와 같이 작은 데이터 셋을 복사하여 큰 데이터 셋과의 비율을 맞추어주는 방법이다. 본 연구 데이터에서는 일반과 불법의 비율에 따라 약 20배 비율에 맞도록 확장하였다.

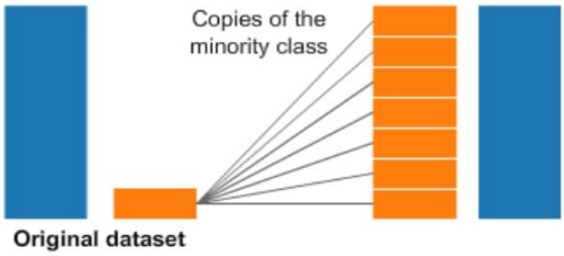

Fig. 2. Oversampling example

3) Machine Learning model

- SVM (Support Vector Machine)

데이터가 분포하고 있는 벡터공간보다 더 낮은 차원인 초평면을 기준으로 데이터를 분류하는 모델이며, [Fig 3]과 같이 모델 학습 시 초평면에서 가장 가까운 데이터의 거리 (마진값)를 최대로 하는 초평면을 구한다. Logistic 분류기에 비하여, 분류 기준에 더 균형적이고, outlier에 강하다는 장점이 있어 보통 더 좋은 성능을 보인다.

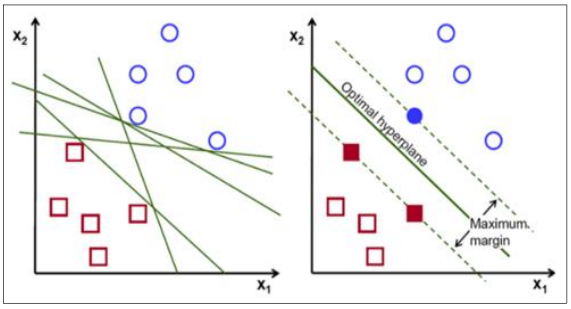

Fig. 3. SVM (Support Vector machine)

데이터에 따라 kernel 함수로 [Fig 4]와 같이 비선형 함수(polynomial kernel, RBF 등)를 사용할 수 있지만, 본 학습 환경과 같이 데이터 집합의 크기가 작은 small dataset의 경우에는 높은 성능을 기대하기 어렵다.

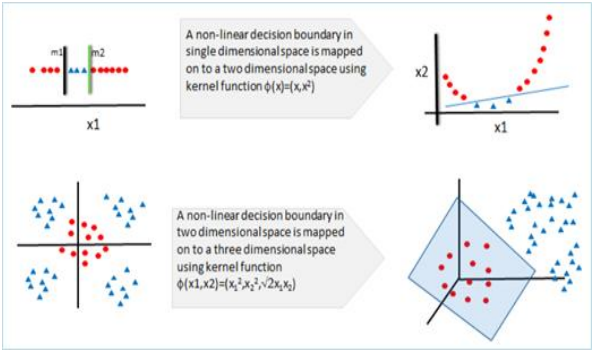

Fig. 4. Kernel function application example

- Random Forest

아래 [Fig 5]와 같이 여러 개의 DT(Decision Tree)로 이루어진 모델로 각각의 DT는 피처와 데이터를 랜덤으로 선택하여 구성된다. 수직/수평적 구조가 아닌 데이터에 낮은 성능을 보인다는 DT의 단점을 보완할 수 있고, 여러 모델의 성능을 합산하는 Ensemble 효과(과적합 방지)를 가진다는 점에서 높은 성능을 보이지만 small dataset 에서는 높은 성능을 기대하기 어렵다.

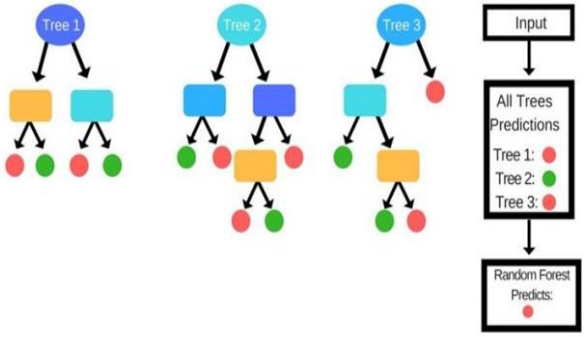

Fig. 5. Random Forest algorithm application example

- CNN(Convolutional Neural Network)

CNN을 문서 분류에 적용하기 위해서는 먼저 Word2Vec과 같은 임베딩 모델의 학습이 필요하다. 본 연구에서는 Word2Vec 모델 학습에 gensim 라이브러리에서 제공하는 임베딩 모델을 사용하였다. 문서 분류에 사용되는 CNN 모델은 각 키워드 단위의 임베딩을 적용한 후, 앞서 언급한 *-gram과 같이 연속형 키워드에 대한 convolution을 적용한 후 일반적인 CNN과 같이 Max-pooling을 적용한다. [Fig 6]은 문서 분류에 사용되

는 CNN 모델의 가장 기본적인 구조이며, 실제 학습에 사용된 CNN은 더 많은 convolution filter를 사용하였다. 그리고 딥러닝의 경우 학습 과정에서 비선형 특성을 검출하거나 학습의 효율을 증가시키기 위하여 activation 함수를 사용하는데, 보통의 경우에 사용하는 relu 함수를 대신하여 sigmoid 함수를 activation 함수로 사용하였다. 이것은 현재 사용하는 데이터셋이 small dataset 이기 때문에, deep 네트워크에 최적화된 relu 함수를 사용하기 어렵기 때문이다.

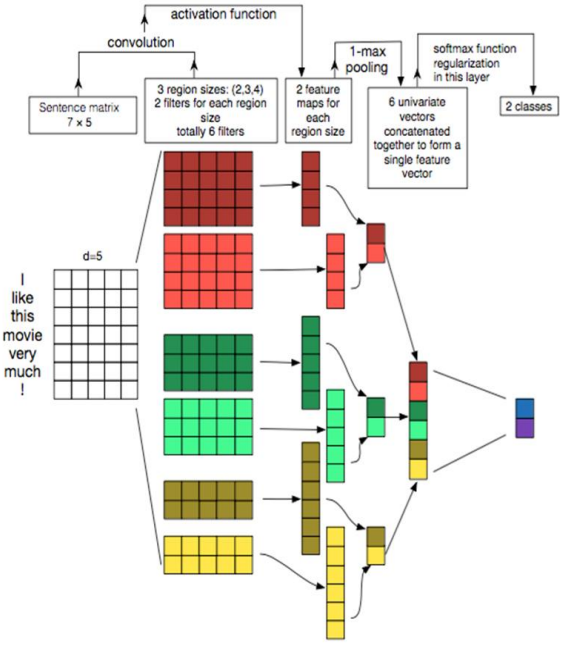

Fig. 6. CNN models example for document classification

III. Classification Model for Illegal Debt Collection

1. Proposed classification model and structure

대부업자의 채권추심 녹취파일을 대상으로 [Fig 7]과 같이 음성을 텍스트로 변환하고 규칙기반 및 기계학습을 통한 분석을 통해 언어폭력, 반복추심 등 위규 혐의를 분석하는 분류 모델을 제안한다. 위규 혐의는 협박, 공포심·불안감 유발, 경조사·병상 중 추심, 대출·불법변제 강요 등 채권추심업무 가이드라인 상 금지행위를 기준으로 하였다.

Fig 7. Classification model for illegal debt collection

① 채권추심 녹취파일을 텍스트로 변환하고 문장을 분석하여 키워드들을 추출

② 추출된 키워드들을 불법추심 판별규칙과 대조하여 위규 혐의를 추출하고, 기계학습 분석 결과에 따른 불법여부 판단

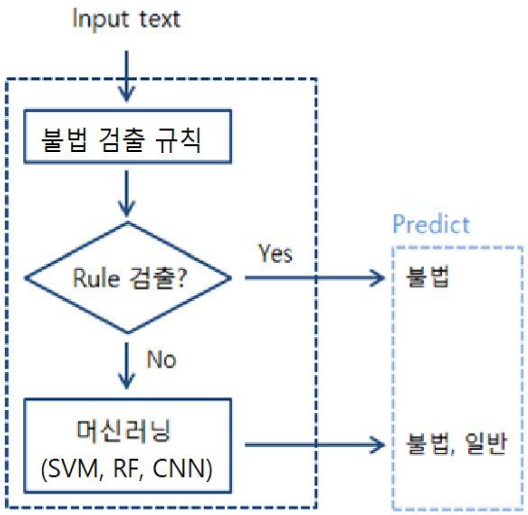

Fig. 8. Structure of the classification model for illegal debt collection

[Fig 8]과 같이 제안한 분류 모델은 규칙기반 불법 검출과 기계학습이 있는데, 불법검출 규칙은 정확히 검출 가능한 불법만을 대상으로 하였기 때문에 사용된 규칙이 검출되지 않은 데이터만을 대상으로 기계학습을 적용하였다. 불법검출 규칙에는 약 80개의 규칙을 적용하였으며, 기계학습에는 학습에 사용되는 데이터의 수가 많지 않기 때문에, 다양한 딥러닝을 적용하기에는 어려움이 있으므로 불법 추심 데이터 분석에는 SVM, RF, CNN을 사용하였다.

불법 규칙과는 다르게 기계학습은 전처리 방법, 알고리즘의 종류, 학습 시점 및 파라미터 튜닝 방법 등에 따라 성능이 달라진다.

2. Data Organization and Preprocessing

1) Data Organization

불법추심 데이터로 학습 및 테스트로 약 1만개의 녹취파일 데이터를 사용하였다. 동 데이터는 대부업 상시감시시스템 구축을 위한 프로젝트 진행을 위해 국내에서 영업을 하고 있는 주요 대부업체 2개사의 협조를 얻어 2019년도 채권추심 녹취 파일 약 9천 여개와 금융당국에서 불법으로 분류한 약 5백 개의 채권 추심 녹취파일을 대상으로 테스트를 실시하였다. 테스트 데이터에 대한 불법 여부 판단은 「채권추심의 공정화에 관한 법률」 등 관련 법률에 따라 실제 실무자가 명확하게 불법으로 판단한 자료를 사용하였다. 음성 데이터를 텍스트 데이터로 만들어주는 과정에서 STT (Speech To Text) 학습을 하였으며, 최종 무작위 추출하여 테스트에 사용된 데이터는 STT 학습을 적용하지 않았다. 학습 및 테스트 데이터의 비율은 아래와 같다.

. 학습 : 일반 8,785개, 불법 388개
. 테스트 데이터 : 일반 80개, 불법 56개
. 테스트 데이터 샘플링 : 일반 50개, 불법 50개 10회 추출 (테스트 데이터 대상 무작위 50:50 추출)

이 때, 테스트 데이터는 기계학습의 성능을 비교하는데 사용되었으며, 샘플링 데이터는 일반, 불법 데이터의 비율을 1:1로 유지한 경우로 최종 선택된 기계학습 기법의 성능을 평가하는데 사용하였다. 테스트 데이터 내의 불법 데이터는 검토를 통해 모호한 데이터는 전부 추출된 상태로 80개 중 56개가 정확한 불법에 해당하였다. 본 연구에서는 '테스트 데이터' (일반 80개, 불법 56개)를 대상으로 각 기계학습 알고리즘(SVM, RF, CNN)별 최적의 모델을 구한 뒤 성능을 비교하고 이 중 가장 높은 성능을 보이는 우수 모델을 선별하고 '테스트 데이터 샘플링'(10회의 샘플 데이터를 무작위 추출)을 대상으로 분류 모델에 따른 성능 분석 실시하고 기계학습만을 적용 했을 때와 분류 정확도 차이를 분석 하였다.

2) Text Data Preprocessing

학습용 데이터셋을 구축하기 위해 텍스트 데이터를 형태소로 분석하는 과정이 필요하며 텍스트 데이터 전처리 방법으로는 POS(Part of Speech) 태깅, 키워드 청킹 등 다양한 방법이 있지만, POS 태깅이 높은 성능을 보여 해당 방법을 사용하였다. POS 태깅은 자연어처리를 통해 문장을 품사가 표시되는 키워드의 형태로 제공하는 것을 말한다. 규칙(rule) 분석, 기계학습 분석 등 분석에 사용되는 방법에 따라 특정 품사만을 사용한다거나, 품사 정보를 제외하고 키워드만을 사용하는 경우도 있다. 성능분석 과정에서는 크게 아래의 세 가지 경우 (pos 태깅, uni-gram, tri-gram)을 가정하였다.

- POS 태깅 : 키워드/품사 형태로 변환 후 사용
- uni-gram : 명사, 형용사, 동사, 부사 품사에 대해 키워드만 사용
- tri-gram : uni-gram 결과 중 인접 키워드 조합을 고려 (3개까지)

Table 2. Preprocessing result example

Original Text	네 안태수 고객 되 나요 리드 코프 에 주효린인데요
POS Tagging	네/AN 안태수/NN 고객/NN 님/SF 되/VB 시/PE 나요/EN 예/EX 리드/NN 코프/NN 에/PP 주효린인데요/NN
uni-gram	네 안태수 고객 되 나요 리드 코프 에 주효린인데요
tri-gram	안태수고객되 고객되리드 되리드코프 리드코프주효린인데요

uni-gram, tri-gram의 경우, 명사, 형용사, 부사, 동사 품사만을 사용하는 이유는 해당 키워드가 가장 많은 의미를 내포하고 있으며, 모든 품사를 사용하는 경우에는 의미 없는 키워드가 많아지면서 주요 피처를 학습하기 어려워지는 문제점이 있다.

3. Rule-Based Illegality Detection Rules

불법 데이터의 개수가 많지 않기 때문에 정확한 불법을 기계학습 이전에 검출할 수 있는 규칙을 사전에 정의하는 것이 필요하다. 예를 들어, 욕설의 경우에는 우리가 사전에 욕설 관련 키워드를 정의하여 검출할 수 있기 때문에, 규칙을 통해 검출이 가능하다. 하지만, 규칙기반 검출을 제외하고 기계학습만을 이용할 시에는 학습 데이터에 다수의 욕설 표현이 등록되어 있지 않거나, 분류가 지식 구축 과정에서 오 분류된 경우가 존재하는 경우 학습된 분류기가 욕설 관련 분류를 완벽하게 수행할 수 없다. [Table

3]은 원문 또는 POS 태깅 결과를 기반으로 하는 규칙 패턴 예시를 나타내었으며, [Table 4]는 규칙 종류별 규칙 예시를 나타내었다. 실제 성능분석에서 사용한 규칙의 개수는 약 80개 정도로 정확히 불법임을 판별할 수 있는 규칙만을 사용하였으며, 이 부분에 대해서는 향후 추가적인 규칙을 생성할 수 있다.

Table 3. Rule pattern example

Rule pattern	Content
str_mat + 띄어쓰기 고려	Original text matching, spaces considered
str_mat + 띄어쓰기 미 고려	Original text matching, spaces not considered
str_mat_pat + 띄어쓰기 미 고려	Original text matching, whether two or more key words are matched, spaces not considered (distance between adjacent key words less than 10)
품사_mat	POS tagging matching, word class + keyword matching
품사_mat_pat_nopos	POS tagging matching, whether two or more key words (not including word class) are matched (the distance between adjacent key words less than 5): type of word class not considered

Table 4. Example of rules according to types

Rule type	Rule pattern	Rule example
언어폭력	str_mat + 띄어쓰기 미고려	새끼, 미친놈,…
언어폭력	품사_mat	시발/NN, 시팔/NN,…
언어폭력	품사_mat_pat_nopos	엿+먹, 사람+무시하,…
경조사(결혼, 장례 등)/병상 중 추심	품사_mat_pat_nopos	병원+입원+하,…
반복 및 야간(오후 9시-아침 12시) 방문	str_mat_pat + 띄어쓰기 미고려	밤늦게+방문,…
제3자 채무사실 고지	품사_mat_pat_nopos	남편+채무+알리,…

IV. Machine Learning Performance Evaluation

SVM, RF, CNN 각각의 기계학습 알고리즘에 대해 최적의 학습기법을 몇 개 선별하여 성능을 확인하고, 이 중 best 학습기법을 선별하여 각 기계학습 알고리즘별 성능을 비교하였다. 이 때, uni-gram의 경우에는 각 기계학습별 3개의 학습기법을 선별하고, tri-gram의 경우에는 2개의 학습기법을 선별하였다. 그리고 CNN의 경우에는 학습 데이터가 small dataset 이기 때문에 초기 파라미터 튜닝이 학습에 거의 영향을 주지 못하여 1개의 학습기법만을 선별하였으며, CNN은 자체에 이전 키워드의 정보를 함께 학습하기 때문에 tri-gram에 대한 성능분석을 진행하지 않았다.

1. Machine Learning Performance Evaluation Method

기계학습 알고리즘별 성능을 평가할 때는 정확도(accuracy) 외에도 정밀도(Precision), 재현율(Recall)에 대해 함께 고려하여야 한다. 향후 실제 기계학습을 적용한 데이터 분석 시에는 정밀도와 재현율이 주요 요소로 사용될 수 있다.

Table 5. Method of organizing classification results

		Classification result	
		True	False
Correct Answer	True	TP	FN
	False	FP	TN
True Positives (TP)	If the correct answer is True, and the model predicts it as True		
False Negatives (FN)	If the correct answer is True, but the model predicts it as False, which is the wrong answer.		
False Positives (FP)	If the correct answer is False, but the model predicts it as True, which is the wrong answer.		
True Negatives (TN)	If the correct answer is False, and the model predicts it as False		

Table 6. Machine learning performance evaluation index

Precision	The proportion of correctly predicting answer to those of predicting that the answer is correct.	$\frac{TP}{TP+FP}$ .....(1), $\frac{TN}{FN+TN}$ .....(2)
Recall	The proportion of correctly predicting the answers to those predicting that actual answers are correct.	$\frac{TP}{TP+FN}$ .....(3), $\frac{TN}{FP+TN}$ .....(4)
Accuracy	Focus on how many of the answers were correct	$\frac{TP+TN}{TP+FN+FP+TN}$ .....(5)

2. Performance Evaluation Result

1) Illegality detection rule

테스트 데이터 대상 불법규칙을 적용하였을 때, 전체 불법 데이터 56개 중 20개의 문서에 대해 불법 규칙이 검출

되었다. 따라서 이후 평가하는 기계학습 모델의 경우, 20개의 문서는 불법으로 검출되었다고 가정하며, 기계학습 입력으로 사용되는 데이터는 일반 80개, 불법 36개가 된다.

2) SVM

[Table 7, 8]은 불법검출 규칙을 통과한 테스트 데이터를 대상으로 식별한 결과이며, [Table 5]를 참고하여 나타내었다.

Table 7. SVM-Uni identification result

Learning Algorithm	Parameters	TN	FP	FN	TP
SVM-Uni-1	loss : log, iter : 3, lr : e^{-2}	46 (Rule:20+ SVM:26)	10	7	73
SVM-Uni-2	loss : huber, iter : 1, lr : e^{-3}	48 (Rule:20+ SVM:28)	8	10	70
SVM-Uni-3	loss : log, iter : 1, lr : $5e^{-2}$	45 (Rule:20+ SVM:25)	11	8	72

Table 8. SVM-Tri identification result

Learning Algorithm	Parameters	TN	FP	FN	TP
SVM-Tri-1	loss : huber, iter : 1, lr : $5e^{-3}$	31 (Rule:20 +SVM:11)	25	1	79
SVM-Tri-2	loss : huber, iter : 3, lr : $5e^{-3}$	29 (Rule:20 +SVM:9)	27	2	78

위 적용 결과를 [Table 6]의 평가 지표를 사용하여 성능분석 결과를 나타내면 아래 [Table 9, 10]과 같다.

Table 9. SVM-Uni performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
SVM-Uni-1	87.95	86.79	91.25	82.14	87.50
SVM-Uni-2	89.74	82.76	87.50	85.71	86.76
SVM-Uni-3	86.75	84.91	90.00	80.36	86.03

Table 10. SVM-Tri performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
SVM-Tri-1	75.96	96.88	98.75	53.36	80.88
SVM-Tri-2	74.29	93.55	97.50	51.79	78.68

3) RF

RF 기계학습에 대한 적용 및 성능분석은 SVM과 동일한 방법으로 진행하였다. [Table 11, 12]는 불법검출 규칙을 통과한 테스트 데이터를 대상으로 적용한 결과이며, [Table 5]를 참고하여 나타내었다.

Table 11. RF-Uni identification result

Learning Algorithm	Parameters	TN	FP	FN	TP
RF-Uni-1	estimators : 20, max depth : 5	38 (Rule:20+ RF:18)	18	7	73
RF-Uni-2	estimators : 30, max depth : 5	32 (Rule:20 + RF:12)	24	4	76
RF-Uni-3	estimators : 50, max depth : 5	30 (Rule:20 + RF:10)	26	5	75

Table 12. RF-Tri identification result

Learning Algorithm	Parameters	TN	FP	FN	TP
RF-Tri-1	estimators : 10, max depth : 10	26 (Rule:20+ RF:6)	30	1	79
RF-Tri-2	estimators : 10, max depth : 15	24 (Rule:20+ RF:4)	32	2	78

위의 적용 결과를 [Table 6]의 평가 지표를 사용하여 성능분석 결과를 나타내면 아래 [Table 13, 14]와 같다.

Table 13. RF-Uni performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
RF-Uni-1	80.22	84.44	91.25	67.86	81.62
RF-Uni-2	76.00	88.89	95.00	57.14	79.41
RF-Uni-3	74.26	85.71	93.75	53.57	77.21

Table 14. RF-Tri performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
RF-Tri-1	72.48	96.30	98.75	46.43	77.21
RF-Tri-2	70.91	92.31	97.50	42.86	75.00

4) CNN

CNN은 앞서 언급한 기계학습 기법과는 다르게 인접 키워드 간 연관관계를 convolution을 이용하여 학습하기 때문에, tri-gram에 대한 성능분석을 수행하지 않았다. 그리고 CNN의 경우 small dataset에 대해서는 파라미터 튜닝

에 따라 성능 차이가 크지 않기 때문에, 사전에 학습을 통해 미리 특정 파라미터 값으로 고정시키고 학습하였다. 식별 및 성능분석 결과는 [Table 15, 16]과 같다

Table 15. CNN-Uni identification result

Learning Algorithm	TN	FP	FN	TP
CNN-Uni	39 (Rule:20+ CNN:19)	17	3	77

Table 16. CNN-Uni performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
CNN-Uni	81.9	92.86	96.25	69.64	85.29

3. Comparison of the Machine Learning Best Algorithm Performances

3개의 기계학습 알고리즘에 대한 성능을 비교하기 위하여, 최적의 파라미터를 적용하여 최고의 성능을 보이는 각 알고리즘의 분류결과 및 성능을 [Table 17, 18]에 나타내었다.

Table 17. Classification of the machine learning best algorithm

Learning Algorithm	TN	FP	FN	TP
SVM Best	46 (Rule:20 + SVM:26)	10	7	73
RF Best	38 (Rule:20 + RF:18)	18	7	73
CNN Best	39 (Rule:20 + CNN:19)	17	3	77

Table 18. Machine learning best algorithm performance

Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
SVM Best	87.95	86.79	91.25	82.14	87.50
RF Best	80.22	84.44	91.25	67.86	81.62
CNN Best	81.91	92.86	96.25	69.64	85.29

3개의 기계학습 모델 중 정확도 측면에서 SVM의 성능이 가장 높음을 확인할 수 있다. 그리고 CNN의 경우, 정확도 측면에서는 낮은 성능을 보이지만 정밀도 측면에서는 높은 성능을 보임을 확인할 수 있다. 이는 CNN이 SVM에 비하여 가지는 이점일 수도 있지만, SVM의 경우 파라미터 튜닝을 통해 CNN 성능과 유사한 성능을 보이는 경우도 존재하였다. 그리고 SVM의 학습 시간이 CNN의 학습 시간에 약 1/20 정도였기 때문에, 성능분석에 사용된 기계학습 알고리즘 중 SVM의 성능이 가장 우수하다고 판단하였다. tri-gram에 대해 각 기계학습 알고리즘별로 따로 언급하지는 않았지만, tri-gram을 적용할 경우 uni-gram에 비하여 전반적으로 높은 정밀도를 가지는 것을 확인할 수 있었다. 그러나 정확도 측면에서 uni-gram에 비해 현저히 낮은 성능을 보이기 때문에, 실제 적용에 대해서는 어떤 부분에 중점을 두고 분석 할 것인지에 따라 추가적 검토가 필요할 것으로 보인다.

4. Results of Performance Analysis Targeting Randomly Extracted Data

테스트 데이터를 대상으로 성능분석을 진행한 결과 SVM의 성능이 가장 높음을 확인하였다. 따라서 본 장에서는 본 연구의 테스트 환경으로 사전에 정의한 테스트 데이터 샘플링에 대해 불법검출 규칙과 최적의 성능을 보인 SVM을 적용하여 성능분석을 하였다. 그리고 해당 데이터를 대상으로 분류모델에서 불법검출 규칙을 제외하고 기계학습만으로 분류한 결과에 대해 분석을 실시하였다. 일반 80개, 불법 56개로 구성된 테스트 데이터를 대상으로 일반 50개, 불법 50개의 데이터를 무작위 추출하는 샘플링을 10회 수행하였으며, 각각의 경우에 대한 판별결과의 평균값을 구하고 성능분석을 해 보았다.

Table 19. Results of applying SVM (Best) to the classification model

Number of Extractions	TN	FP	FN	TP
sample_1	41 (Rule:16+ SVM:25)	9	3	47
sample_2	41 (Rule:20+ SVM:21)	9	3	47
sample_3	42 (Rule:17+ SVM:25)	8	4	46
sample_4	41 (Rule:18+ SVM:23)	9	4	46
sample_5	40 (Rule:19+ SVM:21)	10	3	47
sample_6	40 (Rule:16+ SVM:24)	10	5	45
sample_7	41 (Rule:18+ SVM:23)	9	3	47
sample_8	42 (Rule:20+ SVM:22)	8	5	45
sample_9	42 (Rule:17+ SVM:25)	8	4	46
sample_10	41 (Rule:17+ SVM:24)	9	4	46
average	**41.1**	**8.9**	**3.8**	**46.2**

Table 20. Results of the performance analysis of applying SVM (Best) to the classification model

Machine Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
SVM Best	83.85	91.54	92.40	82.20	87.30

[Table 19, 20] 결과는 10회의 무작위 샘플링 데이터에 대한 성능분석 결과이며, [Table 21, 22] 결과는 분류모델에서 불법검출 규칙을 제외하고 분류한 성능분석 결과이다.

Table 21. Classification results of applying only machine learning SVM (Best) (excluding illegality detection rule)

Number of Extractions	TN	FP	FN	TP
sample_1	39	11	3	47
sample_2	36	14	3	47
sample_3	39	11	4	46
sample_4	38	12	4	46
sample_5	38	12	3	47
sample_6	38	12	5	45
sample_7	37	13	3	47
sample_8	39	11	5	45
sample_9	38	12	4	46
sample_10	37	13	4	46
average	**37.9**	**12.1**	**3.8**	**46.2**

Table 22. Classification performance that applied only machine learning SVM (Best) (excluding illegality detection rule)

Machine Learning Algorithm	Precision		Recall		Accuracy (5)
	(1)	(2)	(3)	(4)	
SVM Best	79.25	90.89	92.40	75.80	84.10

[Table 20, 22]에서 나타나는 것처럼 불법검출 규칙이 함께 적용된 분류모델이 기계학습만을 적용한 경우 보다 정확도가 약 3.2% 우수한 것으로 나타났다. 이는 규칙기반 검출을 제외하고 기계학습만을 이용하는 경우에는 학습 데이터에 다수의 욕설 표현이 등록되어 있지 않거나, 분류가 지식 구축 과정에서 오 분류된 경우가 존재하여 학습된 분류기가 욕설 관련 분류를 완벽하게 수행할 수 없기 때문이다. 향후 규칙 모델에서 사용되는 규칙의 개수가 많지 않기 때문에 이를 추가한다면 추가적인 성능 향상을 기대할 수 있을 것으로 예상된다.

V. Conclusions

본 연구에서는 규칙기반 불법검출 규칙과 기계학습을 결합한 불법 채권추심 분류 모델을 제안하였고, 구축된 모델을 이용해서 사전에 학습하지 않은 데이터를 대상으로 성능 평가 결과에 대해 분석하였다. 불법검출 규칙은 약 80개의 규칙을 적용하였고 기계학습은 SVM, RF, CNN을 적용하였다. 이 때 SVM이 복잡도, 정확도 측면에서 가장 좋은 성능을 보였으며, 불법검출 규칙을 결합한 모델이 우수한 정확도를 나타내었다. 본 연구는 첫 번째, 불법채권추심 음성 데이터를 텍스트로 변환하여 불법검출 규칙과 SVM, RF, CNN 등의 기계학습을 함께 적용해 불법채권추심 분류 모델을 처음으로 제안하였다. 둘째, 형태소 분석과 키워드 기반의 규칙 약 80여개를 만들었으며 규칙 기반의 불법검출과 기계학습을 결합한 분류 결과가 기계학습만을 적용한 분류결과 보다 정확도가 우수함을 확인하였다. 본 연구에서 대부업체로부터 입수한 채권추심 녹취파일을 STT로 변환한 텍스트의 화자 구분이 되어있지 않아 발화의 주체가 누구인지 알기 어려웠다. 김주현, 원정임이 진행하였던 불완전판매 판별 모델 연구[14]와 같이 두 화자가 존재하는 데이터의 특성을 감안하여 화자벡터를 추가하는 방식으로 발화주체에 대한 정보를 모델이 학습할 수 있다면 좀 더 높은 정확도를 보일 것이다. 아울러, 데이터불균형문제(data imbalance)에 대한 효과적인 해결책이 필요하다. 데이터불균형문제는 사기적발 시스템 등 현실의 분류 문제에서 빈번히 등장한다. 각 산업분야의 데이터불균형문제를 해결하기 위한 다양한 연구가 존재하지만, 아직 불법채권추심분류 분야에서 해당 연구가 논의된 적은 없다.[15, 16] 본 연구에서도 역시 학습 데이터의 정상채권추심과 불법채권추심의 비율은 22:1 정도로 정상채권추심에 비해 그 수가 부족하다. 이를 해결하기 위해 오버샘플링 방법을 사용하여 문제를 해결하였지만 불법채권추심판별 분야의 데이터불균형문제를 해결하기에 적정한 방법인지에 대한 연구가 필요하다. 기계학습에서는 SVM이 복잡도, 정확도 측면에서 가장 좋은 성능을 보였고 불법검출 규칙을 함께 적용한 모델이 기계학습만을 적용한 경우보다 정확도가 우수한 것으로 나타났다. 본 연구는 규칙기반 불법검출과 기계학습을 결합한 모델을 통해 불법여부를 분류한 최초의 시도이며 후행연구를 진행하여 모델의 완성도를 높인다면 불법채권 추심행위에 대한 소비자 피해 예방에 크게 기여할 수 있을 것이다.

REFERENCES

[1] Financial supervisory service press release, “Operation Performance of the Financial Supervisory Service 「Illegal Private Financial Reporting Center」 in the first half of the year 19”, 2019.9.17.

[2] Chen, L., Guo, G., & Wang, K. (2011). Class-dependent projection based method for text categorization. Pattern Recognition Letters, 32, 1493-1501.

[3] Ikonomakis, M., Sotiris Kotsiantis, and V. Tampakas. "Text classification using machine learning techniques." WSEAS transactions on computers 4.8 (2005): 966-974.

[4] Azam, Nouman, and JingTao Yao. "Comparison of term frequency and document frequency based feature selection metrics in text categorization." Expert Systems with Applications 39.5 (2012): 4760-4768.

[5] Pranckevičius, Tomas, and Virginijus Marcinkevičius. "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification." Baltic Journal of Modern Computing 5.2 (2017): 221.

[6] Sun, Aixin, Ee-Peng Lim, and Ying Liu. "On strategies for imbalanced text classification using SVM: A comparative study." Decision Support Systems 48.1 (2009): 191-201.

[7] Li, Chenbin, Guohua Zhan, and Zhihua Li. "News text classification based on improved Bi-LSTM-CNN." 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2018.

[8] Kumar, Sudhanshu, Mahendra Yadava, and Partha Pratim Roy. "Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction." Information Fusion 52 (2019): 41-52.

[9] Gräbner, Dietmar, et al. "Classification of customer reviews based on sentiment analysis." ENTER. 2012.

[10] Won-Kyung Lee, Min-Ju Lee, DongSu Seo, “Application of Machine Learning Techniques for the Classification of Source Code Vulnerability”, Korea Institute Of Information Security And Cryptology, Vol 30, pp. 735-743, 2020.8.

[11] Sukjae Choi, Jungwon Lee, Ohbyung Kwon, "Financial Fraud Detection using Text Mining Analysis against Municipal Cyber criminality" Korea Intelligent Information Systems Society 23.3 (2017): 119-138.

[12] Wei, Fusheng, et al. "Empirical study of deep learning for text classification in legal document review." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.

[13] Li, Penghua, et al. "Law text classification using semi-supervised convolutional neural networks." 2018 Chinese Control and Decision Conference (CCDC). IEEE, 2018.

[14] JuHyun Kim, Jung-Im Won, "Discrimination Model On Misselling of Financial Products Using Deep Learning" The Korean Institute of Information Scientists and Engineers, 25.6 (2019): 294-302.

[15] Longadge, Rushi, and Snehalata Dongre. "Class imbalance problem in data mining review." arXiv preprint arXiv:1305.1707 (2013).

[16] Wei, Wei, et al. "Effective detection of sophisticated online banking fraud on extremely imbalanced data." World Wide Web 16.4 (2013): 449-475.

Authors

Tae-Ho Kim received the M.S. degrees in the Department Information security from Korea University, in 2009. Currently he is a Ph. D. student of Graduate School of Information Security at Korea University.

His current research interests include Information Security Policy, Financial Fraud Detection, Electronic Financial Security.

Jong-In Lim received the BS, MS, and Ph.D. degrees in the Department of Mathematics at Korea University. Seoul in 1980, 1982 and 1986. Currently he is a professor of the Graduate School of Information Security at Korea University. Also, he is a former Special Advisor to the President for National Security, Republic of Korea. His main research interest include National Cyber Security, Cyber Warfare, Convergence Security, Privacy andCryptography.