Journal of the Korea Convergence Society
Vol. 9. No. 11, pp. 15-21, 2018

ISSN 2233-4890
https://doi.org/10.15207/JKCS.2018.9.11.015

NIDS의 비정상 행위 탐지를 위한 단일 클래스 분류성능 평가

서재현
원광대학교 컴퓨터·소프트웨어공학과 조교수

Performance Evaluation of One Class Classification to detect anomalies of NIDS

Jae-Hyun Seo
Division of Computer Science & Engineering, WonKwang University, Assistant Professor

요 약 본 논문에서는 단일 클래스만을 학습하여 네트워크 침입탐지 시스템 상에서 새로운 비정상 행위를 탐지하는 것을 목표로 한다. 분류 성능 평가를 위해 KDD CUP 1999 데이터셋을 사용한다. 단일 클래스 분류는 정상 클래스만을 학습하여 공격 클래스를 분류해내는 비지도 학습 방법 중 하나이다. 비지도 학습의 경우에는 학습에 네거티브 인스턴스를 사용하지 않기 때문에 상대적으로 높은 분류 효율을 내는 것이 어렵다. 하지만, 비지도 학습은 라벨이 없는 데이터를 분류하는데 적합한 장점이 있다. 본 연구에서는 서포트벡터머신 기반의 단일 클래스 분류기와 밀도 추정 기반의 단일 클래스 분류기를 사용한 실험을 통해 기존에 없던 새로운 공격에 대한 탐지를 한다. 밀도 추정 기반의 분류기를 사용한 실험이 상대적으로 더 좋은 성능을 보였고, 신규 공격에 대해 낮은 FPR을 유지하면서도 약 96%의 탐지율을 보인다.

주제어 : 침입탐지, 단일 클래스 분류, 비지도 학습, 기계학습, 인공지능

Abstract In this study, we try to detect anomalies on the network intrusion detection system by learning only one class. We use KDD CUP 1999 dataset, an intrusion detection dataset, which is used to evaluate classification performance. One class classification is one of unsupervised learning methods that classifies attack class by learning only normal class. When using unsupervised learning, it difficult to achieve relatively high classification efficiency because it does not use negative instances for learning. However, unsupervised learning has the advantage for classifying unlabeled data. In this study, we use one class classifiers based on support vector machines and density estimation to detect new unknown attacks. The test using the classifier based on density estimation has shown relatively better performance and has a detection rate of about 96% while maintaining a low FPR for the new attacks.

Key Words : Intrusion detection, one class classification, unsupervised learning, machine learning, artificial intelligence

1. 서론

네트워크 기반 침입탐지 시스템 (network-based intrusion detection system) [1,2]은 기존의 공격유형에 대한 서명(signatures)을 기반으로 새로운 침입을 탐지한다. 기존에는 새로운 공격유형이 발견되면 작업자가 직접 새로운 규칙을 추가하는 방식을 사용했다. 최근에는 더 많고 다양한 새로운 공격 유형이 발생함에 따라 작업자가 즉시 대응하기 어려운 측면이 있다. 근래에는 새로운 공격 유형에 대응하기 위해 기존의 서명 기반 방법과

*This paper was supported by Wonkwang University in 2018.
*Corresponding Author : Jae-Hyun Seo (delphia7@wku.ac.kr)
Received September 10, 2018 Revised October 19, 2018
Accepted November 20, 2018 Published November 28, 2018

인공지능 기술을 함께 사용하는 추세이다. 새로운 공격 유형은 어떤 클래스로 구분할지에 대한 기준이 모호하여 기존 데이터를 사용한 명확한 클래스 분류가 어렵다. 기존에 없던 새로운 클래스를 판별하기 위한 방법으로 단일 클래스 분류가 적합하며 이에 대한 많은 연구들[3-8]이 있어왔다. 본 연구에서는 비지도 학습 (unsupervised learning)[9] 중 단일 클래스 분류 (one class classification)[10]를 사용하여 정상 클래스 (normal class)와 공격 클래스 (attack class)를 효과적으로 구분하는 방법을 제안한다.

2장에서는 단일 클래스 분류에 관한 다양한 연구를 살펴보고, 3장에서는 실험 데이터셋을 다룬다. 4장에서 실험 및 결과 분석을 하고, 5장에서 결론 및 향후 연구를 다룬다.

2. 관련연구

이 장에서는 단일 클래스 분류 기법을 사용한 비정상 행위 탐지[11, 12]에 관한 연구를 다룬다. 실험 결과 분석에서 관련 연구와 제안 연구의 장·단점을 다루고자 한다.

Moya와 Hush[13]는 단일 클래스 분류를 위한 폐쇄형 과대 타원(closed hyper-ellipsoidal) 결정 경계를 형성하는 다목적(multiple objective) 학습 알고리즘으로 제한된 이차 네트워크를 소개한다. 네트워크 아키텍처에는 각 결정 경계의 크기, 모양, 위치 및 방향에 대한 독립적인 제어를 제공하는 제한된 제약이 있다. 학습 알고리즘과 함께 아키텍처는 폐쇄형 과대 타원 결정 경계에 대해 양의 고유값(eigenvalues)의 형성을 보장한다. 학습 알고리즘에는 분류 매핑 오류를 최소화하기 위한 방법과 결정 경계의 크기를 최소화하려는 방법이 있다. 저자는 개별적인 기준의 덧셈 조합과 곱셈 조합을 모두 고려하여 경계가 있고 정규화된 개별 목적의 기능적 형태를 선택하는 경험적 증거를 제시한다. 결과적으로 여러 객관적인 기준은 훈련 집합에서 목표가 아닌 패턴의 사용을 요구하지 않고 클래스 내 일반화 및 클래스 밖의 일반화를 달성하기 위해 결정 경계가 필요한 크기만큼 증가 또는 감소 할 수 있게 한다. 결과 네트워크는 대상 데이터만으로 학습할 때 조밀한 폐쇄형 결정 경계를 학습한다. 이 접근법의 장점은 단일 클래스 일반화를 위한 고유의 능력, 목표가 아닌 클래스를 특성화하는 것으로부터 자유로움, 다중 클래스에 대해 폐쇄형 결정 경계를 형성하는 능력을 포함한다.

Tax[14]는 단일 클래스 분류를 위해 여러 모델을 제안하였다. 대부분의 방법은 이상치 검출에 초점을 맞춘다. 개념적으로 이상치 검출을 위한 가장 간단한 방법은 목표 세트 주변의 이상치 데이터를 생성하는 것이다. 그런 다음 분류기가 대상 데이터와 특이점을 구별하도록 훈련시킨다. 제안 논문에서 예로 드는 자동 표적 인식 시스템에서는 물체를 탐지하기 위해 ART-2A와 다중 계층 퍼셉트론을 사용했다. 이 방법은 목표에 가까운 데이터를 생성하지 않는 경우에는 매우 저조한 분류 성능을 보인다. 저자는 훈련을 위한 인공적인 특이점을 만들지 않고 단일 클래스 분류의 해결 방법에 초점을 맞춘다.

Hempstalk 등[15]은 참조 분포 (reference distribution)를 형성하는데 사용되는 밀도 추정기 (density estimator)의 적용과 클래스 확률 추정 (class probability estimation)을 위한 표준 모델의 유도를 결합하여 단일 클래스 분류를 하는 방법을 제안한다. 이 방법에서 참조 분포는 두 번째 인공 클래스를 형성하는데 필요한 인공 데이터를 생성하는데 사용된다. 단일 클래스 분류는 목표 클래스와 이 인공 클래스를 사용하는 다중 클래스 분류 문제로 변경된다. 저자는 참조 분포의 밀도 함수가 이러한 방식으로 얻어진 클래스 확률 추정치와 결합되어 목표 클래스의 밀도 함수의 조정된 추정치를 형성하는 방법을 설명한다. 타이피스트 인식(typist recognition) 문제에서 나온 데이터인 UCI 데이터셋에 대해 제안 방법을 적용하여 우수한 분류 성능을 보였다.

Khan과 Madden[16]은 훈련 데이터의 가용성, 사용된 알고리즘 및 적용된 응용 영역에 기반 한 단일클래스 분류 문제에 대한 분류법을 제시한다. 저자는 제안하는 분류의 각 카테고리를 상세히 다루고, 단일 클래스 분류 알고리즘, 기술 및 방법론에 대한 중요성, 한계 및 응용에 초점을 맞춘다.

Nader 등[17]은 일류 분류 알고리즘을 사용하여 SCADA 시스템에서 침입 탐지를 위한 기계 학습을 사용하는 방법을 연구했다. SCADA (supervisory control and data acquisition) 시스템에서 정보 및 통신 기술의 방대한 사용은 SCADA 네트워크에 의존하는 핵심 인프라에 대한 사이버 공격이 가능하게 한다. 이러한 시스템의 다양한 취약점과 사이버 공격의 이질성으로 인해 기존 침입 탐지 시스템 (IDS)에서는 탐지가 매우 어려워졌

다. 사이버 공격에 대한 모델링이 거의 불가능해졌고 이에 대한 잠재적인 결과는 매우 심각할 수 있다. 이 작업의 주된 목적은 일단 기존 IDS 및 방화벽을 우회하는 악의적인 침입을 탐지하는 것이다. 저자는 단일 클래스 분류를 위해 SVDD (support vector data description)와 커널 주성분 분석 방법을 연구했다. 방사형 기본 함수 (RBF) 커널의 lp-norms에 대해 깊이 있는 연구를 하였다. 이 커널에서 대역폭 매개 변수의 최적의 선택을 찾기 위해 휴리스틱 방법이 제안된다. 실험은 여러 종류의 사이버 공격으로 실제 데이터를 사용하여 수행된다.

3장에서는 KDD 1999 데이터셋에 대한 소개를 하고, 4장에서는 단일 클래스 SVM (one-class SVM) 및 밀도 기반의 단일 클래스 분류기[15]와 성능 비교를 한다. 5장에서는 결론을 도출한다.

3. 데이터셋

KDD CUP 1999 침입탐지 평가 데이터셋[18]을 사용한다. 라벨이 없는 새로운 공격 탐지를 목표로 한다. 사용 알고리즘은 단일 클래스 SVM을 사용한다. 정상 및 공격 트래픽 두 개의 클래스를 사용한다. data_10_percent.gz 파일을 훈련 데이터로 사용하고 corrected.gz 파일을 테스트 데이터로 사용한다. 두 파일은 MySQL 데이터베이스에 입력 후 데이터전처리에 사용된다.

Table 1. Attack types of KDD 1999 dataset

Labels	Attacks	Labels	Attacks
1	normal.	16	satan.
2	buffer_overflow.	17	phf.
3	loadmodule.	18	nmap.
4	perl.	19	multihop.
5	neptune.	20	warezmaster.
6	smurf.	21	warezclient.
7	guess_passwd.	22	spy.
8	pod.	23	rootkit.
9	teardrop.	24	snmpgetattack.
10	portsweep.	25	named.
11	ipsweep.	26	xlock.
12	land.	27	xsnoop.
13	ftp_write.	28	sendmail.
14	back.	29	saint.
15	imap.		

Table1은 KDD 1999 데이터셋에 있는 모든 공격 유형을 나타낸다. Table 2는 훈련 데이터셋의 공격유형으로 라벨이 23이하인 공격 유형만 있다. Table 3은 테스트 데이터셋의 공격유형으로 라벨이 24이상인 신규 공격 유형이 포함되어 있다. 제안 기법에서는 기존 공격유형 및 신규 공격유형을 구분하여 탐지 성능을 도출한다.

Table 2. Attack types of training dataset

Labels	Attacks	Labels	Attacks
1	97,278	13	8
2	30	14	2,203
3	9	15	12
4	3	16	1,589
5	107,201	17	4
6	280,790	18	231
7	53	19	7
8	264	20	20
9	979	21	1,020
10	1,040	22	2
11	1,247	23	10
12	21		

Table 3. Attack types of test dataset

Labels	Attacks	Labels	Attacks
1	60,593	15	1
2	22	16	1,633
3	2	17	2
4	2	18	84
5	58,001	19	18
6	164,091	20	1,602
7	4,367	23	13
8	87	24	7,741
9	12	25	17
10	354	26	9
11	306	27	4
12	9	28	17
13	3	29	736
14	1,098		

Fig. 1은 제안 방법의 데이터 전처리 및 분류에 대한 단계별 절차를 보인다. 훈련데이터셋은 정상 클래스인 97,278 인스턴스를 사용한다. 테스트 데이터셋은 정상 클래스인 60,593 인스턴스와 공격 클래스인 250,436 인스턴스를 사용한다. 실험의 연산량을 줄이기 위해 훈련 데이터셋과 테스트 데이터셋은 WEKA의 Stratified-Remove[19] 기법을 사용하여 1/10의 크기로 줄였다. StratifiedRemove 기법은 데이터셋의 기존 클래스 분포를 유지하면서 데이터셋의 크기를 줄이는 효과를 갖는다.

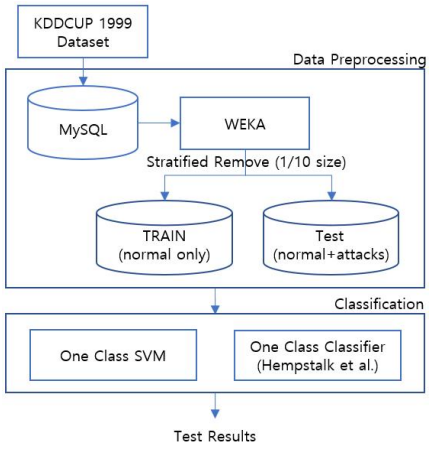

Fig. 1. A flowchart of the proposed method

4. 실험 및 분석

제안 방법은 단일 클래스 SVM 분류기와 Hempstalk 등[15]의 단일 클래스 분류기의 성능을 비교하여 새로운 공격유형에 대한 탐지를 최대화하면서도 기존 공격에 대한 미탐 (false negative) 및 오탐 (false positive)을 최소화하는 것을 목표로 한다.

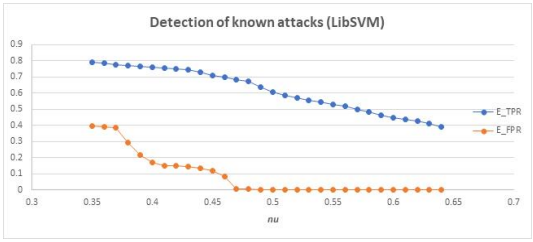

Fig. 2. A graph of TPR and FPR for known attack types

단일 클래스 SVM 실험은 *nu* 및 *gamma* 파라미터를 최적화하는 과정을 거친다. 사전 실험에서 *nu* 파라미터의 범위는 0.01-0.90로 설정하고, *gamma* 파라미터의 범위는 0.1-0.9로 설정하여 최적화를 시도한다. 사전 실험을 통해 파라미터 값의 변화가 실험에 끼치는 영향을 분석한 결과에서 *gamma* 파라미터의 변화는 신규 공격유형 실험 결과에 주요한 영향을 끼치지 않았다.

단일 클래스 SVM에서 판별경계(hyperplane)까지의 거리가 가장 짧은 데이터 벡터를 서포트 벡터라고 한다. *nu* 파라미터는 학습 데이터 중 서포트 벡터로 간주할 최소 비율을 정한다. 또한, 이 비율은 오분류(margin error)를 허용할 수 있는 상한이 된다. 즉, *nu* 값의 증가에 따라 훈련 데이터 중 서포트 벡터가 되는 수가 증가한다. 더불어 오분류에 대한 허용치도 증가한다. 따라서, *nu* 파라미터는 학습 모델이 정상 클래스와 공격 클래스를 구분하는 경계를 조절하는데 많은 영향을 주게 된다.

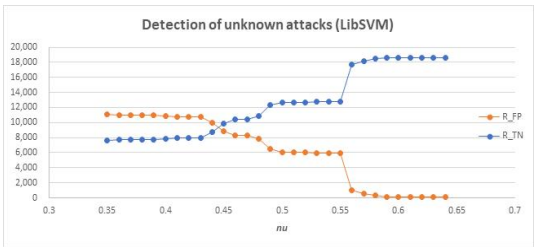

Fig. 3. A graph of FP and TN for new attack types

Fig. 2는 기존 공격유형에 대한 TPR 및 FPR을 비교한 그래프로서 *nu* 파라미터의 값이 0.45 이상의 구간에서 가장 좋은 성능을 보인다. Fig. 3은 FP와 TN 값을 사용하여 새로운 공격유형에 대한 탐지 성능을 보인다. *nu* 값이 0.45보다 작은 경우는 공격 클래스를 정상 클래스로 분류하는 경우가 많았으나 이후 점차 완화되는 현상을 보인다. *nu* 파라미터의 값이 0.55에서 0.6 구간인 경우에 많은 급격한 변화를 보인다. 이 지점부터 공격에 대한 오분류가 낮아진다. 하지만, 동시에 TPR 수치가 낮아지는 결과를 보인다.

Table 4. Confusion matrix [20]

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$TPR(recall) = \frac{TP}{TP+FN} \quad (1)$$

$$FPR(specificity) = \frac{FP}{FP+TN} \quad (2)$$

$$FDS(precision) = \frac{TP}{TP+FP} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$F1\,score = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

Table 5. The measures used in the proposed method

	Measure	Description
Results for normal and all attack types	TPR	Recall
	FPR	Specificity
	ACC	Accuracy
	F1	F1 score
Results for existing attack types	E_TPR	Recall
	E_FPR	Specificity
	E_ACC	Accuracy
	E_F1	F1 score
Results for new attack types	R_FPR	Specificity
	R_ACC	Accuracy

Table 6. TPR, FPR, accuracy, F1-score of one-class SVM according to nu parameter.

nu	TPR	FPR	ACC	F1	R_FPR	R_ACC
0.45	0.711	0.145	0.827	0.615	0.476	0.524
0.46	0.699	0.110	0.853	0.650	0.444	0.556
0.47	0.685	0.037	0.909	0.745	0.442	0.558
...						
0.54	0.544	0.026	0.890	0.658	0.320	0.680
0.55	***0.531***	***0.026***	***0.888***	***0.648***	***0.318***	***0.682***
0.56	*0.520*	*0.006*	*0.901*	*0.673*	*0.053*	*0.947*
0.57	0.498	0.004	0.899	0.657	0.030	0.970
0.58	0.481	0.003	0.897	0.645	0.016	0.984

Table 7. Results of one-class SVM according to nu parameter.

nu	E_TPR	E_FPR	E_ACC	E_F1	R_FPR	R_ACC
0.45	0.711	0.119	0.846	0.657	0.476	0.524
0.46	0.699	0.083	0.872	0.694	0.444	0.556
0.47	0.685	0.004	0.931	0.805	0.442	0.558
...						
0.54	0.544	0.003	0.903	0.700	0.320	0.680
0.55	***0.531***	***0.003***	***0.901***	***0.689***	***0.318***	***0.682***
0.56	***0.520***	***0.002***	***0.899***	***0.680***	***0.053***	***0.947***
0.57	0.498	0.002	0.894	0.661	0.030	0.970
0.58	0.481	0.002	0.891	0.647	0.016	0.984

Table 4는 평가지표인 혼동행렬(confusion matrix)[20]을 나타내고 식 1부터 식 5는 성능평가를 위한 측정치를 나타낸다. Table 5는 실험에 사용한 측정치에 대한 상세 내용을 보인다. Table 6과 Table 7에서는 기존 공격유형에 대한 탐지 성능과 신규 공격유형에 대한 탐지 성능을 비교한다. 실험결과에서 LibSVM 기반의 단일 클래스 SVM 실험은 기존 공격유형에 대한 탐지는 좋은 성능을 보이나, 신규 공격에 대한 탐지 성능은 좋지 않은 것을 볼 수 있다. Table 8은 Hempstalk 등[15]의 단일 클래스 분류기를 사용한 실험 결과를 나타낸다. Hempstalk 등의 단일 클래스 분류기는 밀도 추정기(density estimator)와 클래스 확률 추정(class probability estimation)을 결합하는 방법이다.

Table 8. Results of Hempstalk's one-class classifier

TPR	FPR	ACC	F1	R_FPR	R_ACC
0.961	0.090	0.920	0.824	0.090	0.910

다음은 Hempstalk 등[15]의 방법에 대한 식을 보인다. T를 목표 클래스(target class)라고 하자. 목표 클래스에 대한 훈련 데이터는 주어진다. 제안 실험에서는 정상 클래스에 해당한다. A는 인공 클래스(artificial class)라고 하자. 인공 클래스는 참조 분포(reference distribution)를 사용하여 생성한다. X는 하나의 인스턴스를 나타낸다. $P(X|A)$는 참조 분포의 밀도 함수(density function)이다. $P(X|T)$는 목표 클래스에 대한 밀도 함수로서 구하고자 하는 결과이다.

클래스 확률 함수(class probability function)인 $P(T|X)$를 안다고 가정하자. 실제로 학습 데이터로부터 학습한 클래스 확률 추정기(class probability estimator)를 사용하여 이 함수를 평가할 필요가 있다.

다음은 T에 대한 밀도 함수를 계산하는 방법을 보여준다. 클래스 확률 함수인 $P(T|X)$, 참조 분포 밀도 함수인 $P(X|A)$ 및 $P(T)$가 주어졌을 때, $P(X|T)$는 목표 클래스의 한 인스턴스를 관찰하는 사전 확률(prior probability)을 나타낸다. 베이즈 이론을 적용하면,

$$P(T|X) = \frac{P(X|T)P(T)}{P(X)}$$ 이다.

이진 분류(two class classification) 상황에서, X의 확률은 다른 클래스 라벨을 가진 X의 한 인스턴스를 보게 되는 확률이다. 그래서 식은 다음과 같이 바꿀 수 있다.

$$P(T|X) = \frac{P(X|T)P(T)}{P(X|T)P(T)+P(X|A)P(A)}$$

단일 클래스 분류에 사용하길 바라는 목표 클래스에 대한 밀도 함수인 $P(X|T)$에 대해 푼다. 우선 오른쪽의 분모를 왼쪽으로 가져온다.

$$(P(X|T)P(T)+P(X|A)P(A))P(T|X) = P(X|T)P(T)$$

좌변에 있는 곱을 풀고 $P(X|T)$와 관련된 항을 우측으로 보낸다.

$$P(X|T)P(T)P(T|X)+P(X|A)P(A)P(T|X) = P(X|T)P(T)$$

$$P(X|A)P(A)P(T|X) = P(X|T)P(T) - P(X|T)P(T)P(T|X)$$

그런 후에 $P(X|T)$를 추출하고 나머지는 좌변으로 보낸다. 두 변을 교환하고 분모에서 $P(T)$를 추출한다.

$$\frac{P(X|A)P(A)P(T|X)}{P(T)-P(T)P(T|X)} = P(X|T)$$

$$P(X|T) = \frac{P(X|A)P(A)P(T|X)}{P(T)(1-P(T|X))}$$

두 클래스만이 존재하기 때문에, $P(A)$는 $1-P(T)$와 같다. 다시 정리하면,

$$P(X|T) = \frac{(1-P(T))P(T|X)}{P(T)(1-P(T|X))}P(X|A)$$ 이다.

이 식에서 클래스 확률 함수 $P(T|X)$와 목표 클래스 $P(T)$의 사전 확률을 통해 인공 클래스 $P(X|A)$의 밀도는 목표 클래스 $P(X|T)$의 밀도와 연관된다.

실제로, 이 식을 사용하기 위해 $P(X|A)$를 선택하고 그것으로부터 사용자가 특정하는 양만큼의 인공 데이터를 생성한다. 이 데이터에 있는 각 인스턴스의 클래스는 라벨 A이다. 목표 클래스에 대한 훈련 데이터셋 내에 있는 각 인스턴스는 클래스 라벨 T를 갖는다. 그리고나서 라벨이 붙여진 두 데이터셋을 결합한다. 결합데이터셋에서 T에 속하는 인스턴스들의 비율은 $P(T)$의 추정(estimate)이고, $P(T|X)$의 역할을 갖는 클래스 확률 추정기(class probability estimator)를 획득하기 위해 이진 분류 학습 알고리즘을 적용할 수 있다.

Fig. 4는 LibSVM 기반의 단일 클래스 분류기와 Hempstalk 등의 단일 클래스 분류기를 비교한 그래프이다. 단일 클래스 SVM의 *nu* 파라미터를 변화시키는 실험을 하여 신규 공격에 대한 FPR이 Hempstalk 등의 단일 클래스 분류기를 사용한 실험의 FPR과 유사한 경우에 대한 결과를 비교한다. 실험 결과는 Hempstalk 등의 단일 클래스 분류기를 사용한 실험의 성능이 월등함을 보인다. 제안 방법의 Hempstalk 등의 분류 실험은 밀도 기반 방법만을 사용한 경우의 실험 결과이다.

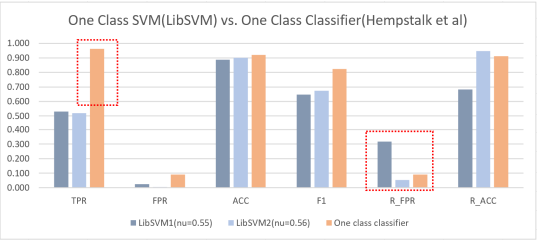

Fig. 4. Performance comparison of one-class SVM (LibSVM) and Hempstalk's one-class classifier

5. 결론

침입탐지 데이터셋에 단일 클래스 분류 기법을 사용하여 기존에 없던 새로운 공격 유형에 대한 탐지를 수행하였다. 분류 성능 평가를 위해 LibSVM 기반의 단일 클래스 분류기와 Hempstalk 등의 단일 클래스 분류기를 사용하였다. TPR 및 FPR 수치를 비교한 실험 결과에서 Hempstalk 등의 단일 클래스 분류기가 상대적으로 우수한 성능을 보였다. 이 분류기는 새로운 공격에 대해 9%의 낮은 FPR을 유지하면서도 약 96%의 탐지 효율을 보인다. 비록, 단일 클래스 분류가 정상 클래스에 대한 학습 데이터만을 사용하므로 좋지 않은 성능을 보일 수 있지만, 새로운 유형의 비정상 행위 탐지를 위해서 단일 클래스 분류의 성능 향상은 중요한 이슈라고 할 수 있다.

향후, 비지도 학습 기반의 딥러닝 방법인 오토인코더(autoencoder)[21,22]와 간스 (GANS, generative adversarial networks)[23] 등을 사용한 단일 클래스 분류 연구를 할 계획이다.

REFERENCES

[1] B. Mukherjee, L. T. Heberlein, & K. N. Levitt. (1994). Network intrusion detection. *IEEE network, 8(3),* 26-41.

[2] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, & E. Vázquez. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security, 28(1-2),* 18-28.

[3] S. S. Khan & M. G. Madden. (2009). A survey of recent trends in one class classification. In *Irish Conference on Artificial Intelligence and Cognitive Science,* 188-197. Springer, Berlin, Heidelberg.

[4] G. Ratsch, S., Mika, B., Scholkopf, & K. R. Muller.

(2002). Constructing boosting algorithms from SVMs: an application to one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9),* 1184-1199.

[5] K. L. Li, H. K. Huang, S. F. Tian, & W. Xu. (2003, November). Improving one-class SVM for anomaly detection. In *Machine Learning and Cybernetics, 2003 International Conference on*, *5*, 3077-3081. IEEE.

[6] G. Giacinto, R. Perdisci, M. Del Rio, & F. Roli. (2008). Intrusion detection in computer networks by a modular ensemble of one-class classifiers. *Information Fusion, 9(1),* 69-82.

[7] I. Kang, M. K. Jeong, & D. Kong. (2012). A differentiated one-class classification method with applications to intrusion detection. *Expert Systems with Applications, 39(4),* 3899-3905.

[8] J. H. Seo. (2018). Detection of Car Hacking Using One Class Classifier. *Journal of the Korea Convergence Society, 9(6)*, 33-38.

[9] L. Portnoy, E. Eskin, & S. Stolfo. (2001). Intrusion detection with unlabeled data using clustering. In *Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001)*.

[10] L. M. Manevitz & M. Yousef. (2001). One-class SVMs for document classification. *Journal of machine Learning research, 2*, 139-154.

[11] J. H. Seo. (2018). Feature Selection for Anomaly Detection Based on Genetic Algorithm, *Journal of the Korea Convergence Society, 9(7)*, 1-7.

[12] J. G. Kang, J. Y. Lee, & Y. Y. You. (2017). A Study on Implementation of Fraud Detection System (FDS) Applying BigData Platform, *Journal of the Korea Convergence Society, 8(4)*, 19-24.

[13] M. M. Moya & D. R. Hush. (1996). Network constraints and multi-objective optimization for one-class classification. *Neural Networks, 9(3)*, 463-474.

[14] D. M. J. Tax. (2001). *One-class classification: concept-learning in the absence of counter-examples* [Ph. D. thesis]. Delft University of Technology, Stevinweg, The Netherlands.

[15] K. Hempstalk, E. Frank, & I. H. Witten. (2008, September). One-class classification by combining density and class probability estimation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 505-519. Springer, Berlin, Heidelberg.

[16] S. S. Khan & M. G. Madden. (2014). One-class classification: taxonomy of study and review of techniques. *The Knowledge Engineering Review, 29(3)*, 345-374.

[17] P. Nader, P. Honeine, & P. Beauseroy. (2014). lp-norms in One-Class Classification for Intrusion Detection in SCADA Systems. *IEEE Transactions on Industrial Informatics, 10(4)*, 2308-2317.

[18] *KDD Cup 1999 Data*, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[19] *WEKA*, https://www.cs.waikato.ac.nz/ml/weka/

[20] *Confusion matrix*, https://en.wikipedia.org/wiki/Confusion_matrix

[21] C. Zhou & R. C. Paffenroth. (2017). Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 665-674.

[22] H. Moeini & F. M. Torab. (2017). Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran. *Journal of Geochemical Exploration, 180*, 15-23.

[23] Y. T. K. Lai, J. S. Hu, Y. H. Tsai, & W. Y. Chiu. (2018). Industrial Anomaly Detection and One-class Classification using Generative Adversarial Networks. In *2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM),* 1444-1449.

서 재 현(Seo, Jae Hyun) [정회원]

- 2008년 2월 : 광운대학교 컴퓨터과학과 (공학석사)
- 2016년 2월 : 광운대학교 컴퓨터과학과 (공학박사)
- 2017년 3월 ~ 현재 : 원광대학교 컴퓨터공학과 교수
- 관심분야 : 최적화, 진화연산, 기계학습
- E-Mail : delphia7@wku.ac.kr