Journal of the Korea Convergence Society
Vol. 9. No. 7, pp. 1-7, 2018

ISSN 2233-4890
https://doi.org/10.15207/JKCS.2018.9.7.001

유전 알고리즘 기반의 비정상 행위 탐지를 위한 특징선택

서재현
원광대학교 컴퓨터·소프트웨어공학과

Feature Selection for Anomaly Detection Based on Genetic Algorithm

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

요 약 데이터 전처리 기법 중 하나인 특징 선택은 대규모 데이터셋을 다루는 다양한 응용분야에서 주요 연구 분야 중 하나로 각광받고 있다. 특징 선택은 패턴 인식, 기계학습 및 데이터 마이닝에서 사용됐고, 최근에는 텍스트 분류, 이미지 검색, 침입 탐지 및 게놈 분석과 같은 다양한 분야에 널리 적용되고 있다. 제안 방법은 메타 휴리스틱 알고리즘 중의 하나인 유전 알고리즘을 기반으로 한다. 특징 부분 집합을 찾는 방법은 크게 필터(filter) 방법과 래퍼(wrapper) 방법이 있는데, 본 연구에서는 최적의 특징 부분 집합을 찾기 위해 실제 분류기를 사용한 평가를 하는 래퍼 방법을 사용한다. 실험에 사용한 훈련 데이터셋은 클래스 불균형이 심하여 희소클래스에 대한 분류 성능을 높이기 어렵다. SMOTE 기법을 적용한 훈련 데이터셋을 사용하여 특징 선택을 하고 다양한 기계학습 알고리즘을 사용하여 선택한 특징들의 성능을 평가한다.

주제어 : 침입탐지, 기계학습, 유전알고리즘, 특징선택, 주성분 분석

Abstract Feature selection, one of data preprocessing techniques, is one of major research areas in many applications dealing with large dataset. It has been used in pattern recognition, machine learning and data mining, and is now widely applied in a variety of fields such as text classification, image retrieval, intrusion detection and genome analysis. The proposed method is based on a genetic algorithm which is one of meta-heuristic algorithms. There are two methods of finding feature subsets: a filter method and a wrapper method. In this study, we use a wrapper method, which evaluates feature subsets using a real classifier, to find an optimal feature subset. The training dataset used in the experiment has a severe class imbalance and it is difficult to improve classification performance for rare classes. After preprocessing the training dataset with SMOTE, we select features and evaluate them with various machine learning algorithms.

Key Words : Intrusion detection, Machine Learning, Genetic Algorithm, Feature Selection, PCA

1. 서론

데이터 마이닝 (data mining) 분야의 데이터 전처리 기법 중 하나인 특징 선택 (feature selection)[1-2]은 중요하고 자주 사용되는 기술이다. 목표로 하는 결과를 도출하는데 있어서 중복 및 잡음(noise) 데이터 등을 제거하여 연산시간을 줄이고 예측 정확도를 높일 수 있다. 일반적으로 특징 선택은 예측 성능 향상 및 예측 시간 단축을 목표로 한다. 대규모 데이터 분석을 필요로 하는 연구 분야에서 특징선택의 중요성은 상당히 높다. 특징 선택은 패턴 인식, 기계 학습 및 데이터 마이닝에서 연구 개발의 중요한 분야였으며 인터넷 문서의 텍스트 분류, 이

*This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP; Ministry of Science, ICT & Future Planning) (No. NRF-2017R1C1B5018128)
*Corresponding Author : Jae-Hyun Seo (delphi7@wku.ac.kr)
Received April 16, 2018 Revised May 3, 2018
Accepted July 20, 2018 Published July 28, 2018

미지 검색, 고객 관계 관리, 침입 탐지 및 게놈 분석과 같은 많은 분야에 널리 적용되어왔다. 본 논문에서는 기존의 다양한 침입탐지 관련 연구들[3-5]을 다양하게 참고하였다.

특징 선택 [1,6]은 초기 특징 집합의 부분 집합을 선택하는 과정으로 특징 서브셋 (feature subset)들은 특정 평가 기준에 의해 측정된다. 일반적으로 최적화된 특징 하위 집합을 찾는 것은 NP-hard [7] 문제로서, 근사 최적해를 찾기 위해 메타 휴리스틱 (meta-heuristic) 기반의 알고리즘을 주로 사용한다. 특징 선택 과정은 부분 집합 생성, 부분 집합 평가, 중지 기준 및 결과 검증의 단계로 이루어진다.

특징 선택 과정에서 특징 부분 집합을 생성 및 평가하는 방법 [8]은 필터 (filter) 방법과 래퍼 (wrapper) 방법으로 구분할 수 있다. 필터 방법은 특징 부분 집합에 대한 평가 기준이 독립적인 방법이기 때문에 어떤 특징을 어떠한 방법으로 평가하는지에 따라 실제 평가 성능이 많이 달라질 수 있다. 이 방법은 특징 부분 집합을 평가하는 시간은 빠르지만, 실제 분류기를 사용할 때 많은 성능 차이를 보일 수 있는 단점이 있다. 래퍼 방법은 특징 부분 집합에 대한 평가를 종속적인 형태로 한다. 이 방법은 실제 분류기를 사용하여 특징 부분 집합을 평가하기 때문에 평가 시간이 오래 걸린다는 단점이 있다. 하지만, 실제 분류시의 성능이 우수한 특징 부분 집합을 찾는데 유리하다. 본 연구에서는 침입탐지 데이터셋을 사용한 래퍼 방법 기반의 특징 선택을 통하여 네트워크 침입 탐지 데이터셋 분류 성능을 높이고 침입탐지 시스템의 성능을 개선하고자 한다.

2장에서는 특징 선택에 관한 연구를 살펴보고, 3장에서는 실험 데이터셋을 다룬다. 4장에서는 실험 및 결과 분석을 한다. 5장에서는 결론 및 향후 연구를 다룬다.

2. 관련연구

이 장에서는 특징선택 기법을 사용한 비정상 행위 탐지에 관한 연구를 다룬다. 실험 결과 분석에서 관련 연구와 제안 연구의 장·단점을 다루고자 한다.

Jain과 Zongker [6]는 Pudil 등이 제안한 SFFS (sequential forward floating selection) 알고리즘[9]이 다른 알고리즘들보다 우수하다는 것을 보인다. 저자는 네 가지의 텍스처 모델을 사용한다. SAR 위성 이미지를 사용하여 토지 이용 분류를 위한 최적의 특징 집합을 도출한다. 다른 텍스처 모델에서 도출된 풀링 특징들 (pooling features)은 높은 분류 정밀도를 보인다.

Bolon-Canedo 등[10]은 이산화기(discretizer), 필터 및 분류기의 조합으로 구성된 새로운 방법을 제시한다. 그 목적은 분류기의 성능을 향상시키면서도 상당히 적은 특징 집합을 사용한다. 이 방법은 이진 및 다중 클래스 분류 문제에 적용된다. KDD CUP 1999 벤치 마크 데이터를 사용하여 효율성을 테스트한다. 다른 방법 및 KDD 수상자의 연구들과 비교 연구를 한다. 도출된 결과는 제안 방법의 타당성을 보여주고, 대부분의 경우에 더 나은 성능을 달성하면서도 특징의 수를 80 % 이상 줄였다.

Nguyen 등[11]은 기계 학습에서 사용되는 필터 (filter) 방법에 기반 한 자동 특징 선택 절차를 제안한다. 특히, CFS (Correlation Feature Selection)에 중점을 둔다. CFS 최적화 문제를 다항식 혼합 0-1 분수 프로그래밍 문제로 변환하고 추가 변수를 도입함으로써 새로운 혼합 0-1 선형 프로그래밍 문제를 만든다. 혼합 0-1 선형 프로그래밍 문제는 분기 한정법 (branch and bound) 알고리즘을 사용하여 해결한다. 특징 선택 알고리즘인 최상우선 기반 CFS와 유전 알고리즘 기반 CFS 방법을 사용한 비교 실험을 한다. KDD CUP 1999 침입탐지 데이터셋에 C4.5 및 베이지안 네트워크를 적용한다. 선택된 특징을사용하여 분류 정확도를 테스트한다. 제안 방법은 많은 잉여 특징들을 제거하면서도 최상 우선 및 유전 알고리즘 기반의 탐색 전략을 능가한다. 또한, 분류 정확도는 동일하거나 더 나은 성능을 보인다.

Chou 등[12]은 침입 탐지 설계에서 두 단계 접근 방식을 제안한다. 첫 번째 단계에서는 원래의 고차원 데이터에서 쓸모없는 정보를 제거하기 위해 상관 기반 특징 선택 알고리즘을 개발한다. 다음으로 제한적이고 모호한 정보로 인한 불확실성 문제를 해결하기 위해 침입 탐지 방법을 설계한다. 실험에서 평가 도구로 6 개의 UCI 데이터베이스와 DARPA KDD 1999 침입 탐지 데이터 세트를 사용한다. 경험적 연구에 따르면 저자의 특징 선택 알고리즘은 데이터 세트의 크기를 줄일 수 있음을 보인다. 저자의 침입 탐지 방법은 다른 침입 탐지 시스템보다 나은 성능을 제공한다. Table 1은 관련 연구를 비교하여 정리한 표이다.

Table 1. Comparison of related works

Authors	Algorithm	Dataset
Jain과 Zongker	SFFS	SAR satellite images
Bolon-Canedo 등	Ensemble	KDD CUP'99
Nguyen 등	CFS, C4.5, BayesNet	KDD CUP'99
Chou 등	CFS	UCI, KDD CUP'99

3. 침입탐지 데이터셋

실험에 사용하는 KDD CUP 1999 침입탐지 데이터셋 [13]은 이다. 이 데이터셋은 제 3 회 국제 지식 발견 및 데이터 마이닝 도구 공모전(The Third International Knowledge Discovery and Data Mining Tools Competition)에 사용된 데이터셋으로 고성능의 네트워크 침입 탐지 시뮬레이션을 목적으로 한다. 이 데이터셋은 군사 네트워크 환경에서 시뮬레이션 된 다양한 공격 유형을 포함한다.

Table 2. Four attack categories [13]

Category	Comments
DoS	denial-of-service, e.g. syn flood;
R2L	unauthorized access from a remote machine, e.g. guessing password;
U2R	unauthorized access to local superuser (root) privileges, e.g., various ''buffer overflow'' attacks;
Probe	surveillance and other probing, e.g., port scanning.

Table 3. Original training dataset

classes	# of instances	ratio(%)
Normal	97,278	8.8918
U2R	52	**0.0048**
R2L	1,126	**0.1029**
DoS	991,458	90.6251
Probe	4,107	**0.3754**

실험을 위해 KDD CUP 1999 침입탐지 데이터셋 중 라벨(label)이 있는 데이터를 사용한다. 훈련 데이터는 kddcup.data.gz의 10%에 해당하는 데이터를 사용하고 테스트 데이터는 corrected.gz를 사용한다. 실험에 사용하는 공격 유형은 두 데이터셋에 모두 있는 24가지 공격으로 한정하고 Table 2의 네 가지 공격 카테고리 중 하나에 속한다.

Table 4. Training dataset after applying SMOTE

classes	# of instances	ratio(%)
Normal	97,278	8.7538
U2R	6,240	0.5615
R2L	10,134	0.9119
DoS	991,458	89.2185
Probe	6,160	0.5543

Table 5. Test dataset

classes	# of instances	ratio(%)
Normal	60,593	20.7297
U2R	39	0.0133
R2L	5,993	2.0503
DoS	223,298	76.3934
Probe	2,377	0.8132

Table 6. Features of KDD 1999 dataset [13]

#	Features	#	Features
1	duration	22	is_guest_login
2	protocol_type	23	count
3	service	24	srv_count
4	flag	25	serror_rate
5	src_bytes	26	srv_serror_rate
6	dst_bytes	27	rerror_rate
7	land	28	srv_rerror_rate
8	wrong_fragment	29	same_srv_rate
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	dst_host_srv_count
13	num_compromised	34	dst_host_same_srv_rate
14	root_shell	35	dst_host_diff_srv_rate
15	su_attempted	36	dst_host_same_src_port_rate
16	num_root	37	dst_host_srv_diff_host_rate
17	num_file_creations	38	dst_host_serror_rate
18	num_shells	39	dst_host_srv_serror_rate
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmds	41	dst_host_srv_rerror_rate
21	is_host_login		

Table 3과 Table 4는 훈련 데이터셋의 클래스별 분포를 나타낸다. Table 3은 원래 훈련 데이터셋의 분포를 나타낸다. Table 4는 SMOTE (Synthetic Minority Over-sampling Technique) [14] 기법을 사용하여 훈련 데이터를 전처리를 한 후의 비율을 나타낸다. Table 3에서 0.5% 이내의 비율을 차지하는 U2R, R2L 및 Probe 클래스를 희소클래스로 정한다. Table 4는 SMOTE를 사용하여 희소 클래스들의 인스턴스(instance) 수를 증가시킨 후의 훈련 데이터셋 비율을 나타낸다. SMOTE [14]는 비율이 낮은 클래스의 데이터를 만들어 내는 방법으로 해당 데이터의 샘플을 취한 뒤, 이 샘플의 k 최근접 이웃(k-nearest neighbor)을 찾고, 현재 샘플과 이들 k개 이웃 간의 차이(difference)를 구하고, 이 차이에 0~1 사이의 임의의 값을 곱하여 만든 새로운 샘플을 훈련 데이터에 추가한다. 결과적으로 SMOTE는 기존의 샘플을 주변의 이웃을 고려해 약간씩 이동시킨 점들을 추가하는 방식으로 동작한다. Table 5는 테스트 데이터셋의 클래스별 비율을 나타낸다. Table 6은 KDD CUP 1999 침입탐지 데이터셋의 특징들을 나타낸다.

4. 실험 및 분석

KDD 1999 침입탐지 데이터셋을 사용하여 희소 클래스인 U2R, R2L 및 Probe 클래스의 성능 개선에 목표를 둔 특징선택을 시도한다. 원래 데이터셋과 SMOTE 전처리한 데이터셋을 사용한다. 불균형 데이터셋을 사용한 분류에서 희소클래스의 인스턴스(instance) 수가 현저히 적은 경우에 정상적인 분류를 하기 어렵다. 대부분 다수 클래스(majority class)로 모든 분류가 이루어지는 경향이 있다. 이러한 클래스 불균형을 완화하기 위해 SMOTE 데이터셋을 사용한 비교를 한다.

가장 범용적인 특징선택 방법인 주성분 분석법 (PCA, Principal Components Analysis)과 유전알고리즘을 사용한 래퍼 특징선택 방법으로 주요 특징들을 선택하고, 선택한 특징을 사용하여 k-NN, 의사결정트리(DT, Decision Tree) 및 SVM 알고리즘을 적용하여 성능을 비교·분석한다. k-NN의 k는 3으로 한다. 특징 분석을 위해 대표적인 데이터마이닝 소프트웨어 중의 하나인 WEKA (Waikato Environment for Knowledge Analysis) [15]를 사용한다.

주성분 분석법을 사용하여 선택한 특징들은 Table 6의 특징들 중 1에서 20 번까지이다. Table 7과 Table 8은 선택한 특징들에 대한 k-NN, 의사결정트리 및 SVM의 실험 결과를 나타낸다. Fig. 1의 실험결과에서와 같이 희소 클래스인 U2R, R2L 및 Probe 클래스의 분류에서 뚜렷하게 어떤 알고리즘이 우수한 성능을 보인다고 판단하기 어렵다. 원래 훈련 데이터셋을 사용한 실험에서는 k-NN의 성능이 우수하나, SMOTE 데이터셋을 사용한 실험에서는 알고리즘별로 희소 클래스에 대한 성능이 제각각으로 주성분 분석법을 사용한 특징 선택에 문제가 있음을 보인다.

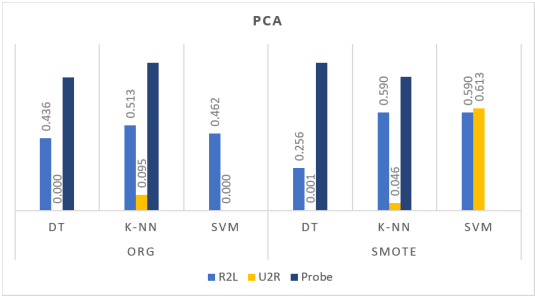

Fig. 1. Results of PCA

Table 7. Recall of PCA with original dataset

Class	k-NN	Decision Tree	SVM
Normal	0.977	0.989	0.991
U2R	**0.513**	0.436	0.462
R2L	**0.095**	0.000	0.000
DoS	1.000	1.000	0.995
Probe	**0.887**	0.802	0.000
w.avg	0.975	0.975	0.965

Table 8. Recall of PCA with SMOTE dataset

Class	k-NN	Decision Tree	SVM
Normal	0.976	0.989	0.990
U2R	**0.590**	0.256	**0.590**
R2L	0.046	0.001	**0.613**
DoS	1.000	1.000	0.994
Probe	0.803	**0.887**	0.000
w.avg	0.974	0.976	0.977

래퍼 특징선택 기법 (Wrapper Subset Evaluation) [8]과 유전 알고리즘 [16]을 사용한 WrapperGA 실험에서는

Table 6의 특징들 중에서 1, 2, 3, 4, 5, 7, 8, 10, 12, 14, 15, 21, 22, 23, 26, 27, 28, 30, 31, 32, 35, 36, 38, 40, 41 에 해당하는 25개 특징들이 선택되었다.

Table 9는 래퍼 특징선택 기법의 파라미터 설정을 나타내고 Table 10은 사용한 유전 알고리즘의 파라미터 설정을 나타낸다. Table 11과 Table 12는 WrapperGA를 사용한 실험 결과로 Table 11은 원래 훈련 데이터셋을 사용한 실험 결과이다. 희소 클래스들에 대한 재현율(recall) 수치에서 의사결정트리가 상대적으로 높은 것을 알 수 있다. Table 12는 SMOTE가 적용된 훈련 데이터셋을 사용한 실험 결과이다. 의사결정트리를 사용한 실험 결과가 다른 실험들에 비해 우수함을 알 수 있다. Table 13은 가장 좋은 결과를 보인 의사결정트리 실험의 혼동행렬을 나타낸다. Fig. 2는 유전알고리즘 기반의 특징 선택 기법을 사용한 실험 결과를 나타낸다.

Table 14와 Table 15는 특징선택을 하지 않고 SMOTE 전처리만 수행한 이전 연구[17]의 실험 결과이다. 이전 연구와 제안 방법의 희소 클래스에 대한 분류 성능 비교에서 U2R은 0.46에서 0.487로, R2L은 0.19에서 0.777로, Probe는 0.92에서 0.997로 각각 성능이 개선되어 유전 알고리즘 기반의 특징선택 실험 결과가 이전 연구에 비해 상당히 개선되었음을 알 수 있다. Fig. 3은 이전 연구[17]의 실험 결과와 비교한 그래프이다. 의사결정트리를 사용한 실험에서 R2L 클래스 탐지 성능이 두드러지게 향상됨을 알 수 있다.

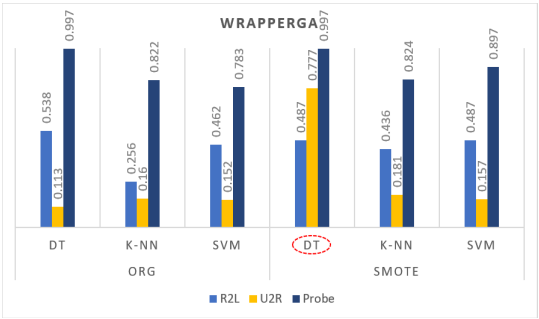

Fig. 2. Results of WrapperGA

Table 9. Parameters of Wrapper Subset Evaluation

Parameter	Value
Classifier	Decision tree
Evaluation measure	Accuracy
Folds	5
Threshold	0.01

Table 10. Parameters of genetic search

Parameter	Value
Population size	20
# of generations	20
Probability of crossover	0.6
Probability of mutation	0.033
Objective function	Accuracy of a classifier

Table 13과 Table 15의 혼동행렬(Confusion matrix)에서 Normal 클래스는 0, U2R은 1, R2L은 2, DoS는 3, Probe는 4로 표기한다.

Table 11. Recall of WrapperGA with original dataset

Class	*k*-NN	Decision Tree	SVM
Normal	0.996	0.983	0.986
U2R	0.256	**0.538**	0.462
R2L	**0.160**	0.113	0.152
DoS	0.997	1.000	0.993
Probe	0.822	**0.997**	0.783
w.avg	0.978	0.978	0.972

Table 12. Recall of WrapperGA with SMOTE dataset

Class	*k*-NN	Decision Tree	SVM
Normal	0.995	0.981	0.982
U2R	0.436	**0.487**	0.487
R2L	0.181	**0.777**	0.157
DoS	0.997	0.996	0.992
Probe	0.824	**0.997**	0.897
w.avg	0.978	0.988	0.972

Table 13. Confusion matrix of WrapperGA with SMOTE dataset (decision tree)

		Actual				
		0	1	2	3	4
Predicted	0	59,448	13	21	435	676
	1	13	**19**	7	0	0
	2	1,271	67	**4,654**	1	0
	3	19	0	0	222,372	907
	4	3	1	0	4	**2,369**

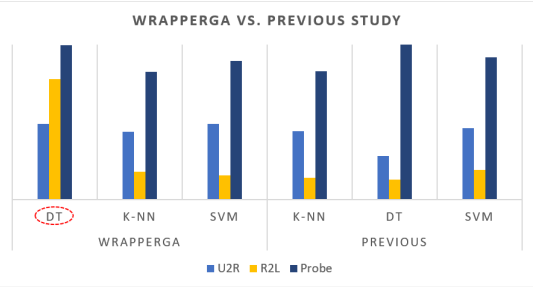

Fig. 3. Comparison of WrapperGA and previous study

Table 14. Recall of previous study (SMOTE dataset) [17]

Class	*k*-NN	Decision Tree	SVM
Normal	1.00	1.00	0.99
U2R	0.44	0.28	**0.46**
R2L	0.14	0.13	**0.19**
DoS	1.00	1.00	0.87
Probe	0.83	1.00	**0.92**

Table 15. Confusion matrix of previous study (SVM with SMOTE dataset) [17]

		Actual				
		0	1	2	3	4
Predicted	0	59,664	82	32	683	132
	1	13	18	7	1	0
	2	4,782	77	1,132	2	0
	3	28,764	0	0	194,521	13
	4	189	0	0	3	2,185

5. 결 론

침입탐지 데이터셋의 희소 클래스에 대한 분류 성능 개선을 위해 주성분 분석법과 유전 알고리즘 기반의 래퍼 특징선택 기법을 적용하여 탐지 성능을 비교 연구하였다. SMOTE 전처리 여부에 따른 실험에서 SMOTE를 적용한 경우에 더 정확한 결과를 도출할 수 있었다. 이전 연구와의 비교에서 제안하는 방법의 탐지 성능이 더 우수하였다. 제안하는 특징선택 기법은 기계학습을 사용한 네트워크 공격, 차량 해킹 및 비정상 행위 탐지 등에 사용 가능할 것으로 보인다. 특히, 네트워크 보안 분야에서 다양한 공격 데이터의 특징을 추출하는데 활용될 것으로 생각한다.

향후, 네트워크 침입 탐지에 대한 효율 개선을 위해 새로운 데이터 전처리 방법 및 특징 선택 기법을 제시하고자 한다.

REFERENCES

[1] H. Liu & L. Yu. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering, 17(4),* 491-502.

[2] I. Guyon & A. Elisseeff. (2003). An introduction to variable and feature selection. *Journal of machine learning research,* 3(Mar), 1157-1182.

[3] E. M. Yang, H. J. Lee & C. H. Seo. (2017). Comparison of Detection Performance of Intrusion Detection System Using Fuzzy and Artificial Neural Network. *Journal of Digital Convergence, 15(6),* 391-398.

[4] H. Y. Lee & H. S. Y. (2014). Quality Evaluation Model for Intrusion Detection System based on Security and Performance. *Journal of Digital Convergence, 12(6),* 289-295.

[5] H. Y. Lee & H. S. Y. (2015). Convergence Performance Evaluation Model for Intrusion Protection System based on CC and ISO Standard. *Journal of Digital Convergence, 13(5),* 251-257.

[6] A. Jain & D. Zongker. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence, 19(2),* 153-158.

[7] A. Blum & R. L. Rivest. (1989). Training a 3-node neural network is NP-complete. In *Advances in neural information processing systems,* 494-501.

[8] R. Kohavi & G. H. John. (1997). Wrappers for feature subset selection. *Artificial intelligence, 97(1-2),* 273-324.

[9] P. Pudil, J. Novovičová & J. Kittler. (1994). Floating search methods in feature selection. *Pattern recognition letters, 15(11),* 1119-1125.

[10] V. Bolon-Canedo, N. Sanchez-Marono & A. Alonso-Betanzos. (2011). Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset. *Expert Systems with Applications, 38(5),* 5947-5957.

[11] H. Nguyen, K. Franke & S. Petrovic. (2010, February). Improving effectiveness of intrusion detection by correlation feature selection. In *Availability, Reliability,*

and Security, 2010. ARES'10 International Conference on, 17-24.

[12] T. S. Chou, K. K. Yen & J. Luo. (2008). Network intrusion detection design using feature selection of soft computing paradigms. *International journal of computational intelligence, 4(3),* 196-208.

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

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16,* 321-357.

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

[16] D. E. Goldberg. (1989). Genetic Algorithms in Search, Optimization & Machine Learning. Addison. *Wesely Publishing Co., Inc, 1998(3),* 25.

[17] J. H. Seo. (2015). A study on the performance evaluation of unbalanced intrusion detection dataset classification based on machine learning. *Journal of the Korean Institute of Intelligence Systems, 27*, 466 - 474.

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