침입탐지시스템에서의 특징 선택에 대한 연구

한 명 묵*

요 약

침입은 컴퓨터 자원의 무결성, 기밀성, 유효성을 저해하고 컴퓨터 시스템의 보안정책을 파괴하는 일련의 행위의 집합이다. 이러한 침입을 탐지하는 침입탐지시스템은 데이터 수집, 데이터의 가공 및 축약, 침입 분석 및 탐지 그리고 보고 및 대응의 4 단계로 구성되어진다. 침입탐지시스템의 방대한 데이터가 수집된 후, 침입을 효율적으로 탐지하기 위해서는 특징 선택이 중요하다. 이 논문에서 유전자 알고리즘과 결정트리를 활용한 특징 선택 방법을 제안한다. 또한 KDD 데이터에서 실험을 통해 방법의 유효성을 검증한다.

A Study for Feature Selection in the Intrusion Detection System

Myung-Mook Han*

ABSTRACT

An intrusion can be defined as any set of actors that attempt to compromise the integrity, confidentiality and availability of computer resource and destroy the security policy of computer system. The Intrusion Detection System that detects the intrusion consists of data collection, data reduction, analysis and detection, and report and response.

It is important for feature selection to detect the intrusion efficiently after collecting the large set of data of Intrusion Detection System. In this paper, the feature selection method using Genetic Algorithm and Decision Tree is proposed. Also the method is verified by the simulation with KDD data.

Key words : Intrusion Detection System, Feature Selection, Genetic Algorithm, Decision Tree

* 경원대학교 소프트웨어대학

1. 서 론

인터넷을 통한 전자상거래, 홈뱅킹 등 네트워크를 이용한 새로운 서비스들이 다양하게 개발되어 사용자들이 급속히 증가하고 있다. 그러나 이를 악용하는 불건전정보 유통 및 정보범죄와 같은 정보화의 역기능 또한 증가하고 있다.

네트워크상에서 발생되는 문제에 대하여 안정적이면서 효율적인 환경을 제공하기 위해서 네트워크상에 존재하는 각종 자원들을 감시할 수 있는 네트워크의 보안 관리(Network Security management)가 필요하게 되었다. 네트워크의 보안 관리를 위해서는 네트워크에서 수집한 패킷을 분석하여 외부로부터의 침입을 사전에 탐지할 수 있는 침입탐지시스템(Intrusion Detection System : IDS)이 필요하다[1].

현재 사용되어지고 있는 침입 탐지를 위한 접근방법에는 두 가지가 있다. 첫 번째 접근방법은 동적인 시스템 사용자의 행위를 and/or의 형태로 특성화시킨 정상적인 패턴 - 보통은 통계적인 방법을 사용하여, 사용자 행위들 간의 관계를 정의하여, 그 정의된 것으로 비정상적인 사용자를 탐지하는 방법이다. 이것을 비정상행위 탐지(anomaly detection)라 한다.

두 번째 접근방법은 오용 탐지(misuse detection) 방법이다. 이것은 이미 알려진 공격방법, 비정상적인 사용자들의 행위, 일정한 코드들의 수행 등을 정형화하여 비정상적인 사용자들을 탐지한다.

지난 40년 동안, 많은 연구들이 특징 선택을 위한 방법들을 개발하는데 노력하여 왔다. 이러한 설계들의 내용은 다양한 특징들의 적절성, 성취성, 그리고 복잡도에 관련이 있다. 특징 선택방법들은 평가를 위해 무엇을 사용했느냐에 따라 필터(filter)와 래퍼(wrapper) 접근 방법으로 나뉜다. 필터 접근 방법은 분류과정 전에 인스턴스(instance)들의 사이를 어떤 거리의 측정치를 바탕으로 특징들의 부분집합을 선택하고, 래퍼 접근 방법은 분류과정 중에 분류의 결과를 바탕으로 선택한다[2~8].

유전자 알고리즘(Genetic Algorithm : GA)은 자연의 적자생존과 진화의 방법을 모방한다. 특별히 최적화 문제를 풀기 위해서 보답(payoff)을 올리거나 값을 내리는 확률적 탐색 기법을 이용한다. 또한 GA는 다중모드(multimodal) 탐색 도메인에서 전역 해를 최적으로 발견하는데 높은 확률을 가지고 있다. 특징 부분집합 선택은 모든 가능한 부분집합 중에 분류 성능을 최대화하는 특징들의 부분집합을 선택하는 과정으로 정의된다. 매우 큰 문제에서는 탐색해야 할 탐색 공간은 매우 크며, 특히 특징 선택 문제는 다중 기준(multicriteria)과 제약(constraint)을 갖는 최적화 문제를 나타낸다. 이러한 점은 GA가 특징 선택 문제에 적용될 수 있다는 것을 보여준다[18].

본 논문에서는 대용량의 IDS 데이터에서 필요한 특징들만을 GA를 활용해서 선택하는 방법을 제안한다. 또한 선택하는 과정에서 결정트리를 활용해서 분류의 성능을 측정한다.

논문은 다음과 같이 구성된다. 2장에서 일반적인 관련 연구에 대해서 설명한다. 3장에서 제안하는 모델인, GA를 활용한 특징 선택 방법을 설명한다. 실험 환경과 결과가 4장에서 설명한다. 마지막으로 결론과 미래 연구 방향이 5장에 서술한다.

2. 관련 연구

2.1 특징 선택 문제

분류과정에서의 첫 번째 단계는 크고 원래의 특징 집합에서 작은 특징의 부분 집합을 선택하는 특징 선택이다. 인스턴스는 분류 알고리즘에 값 $\{v_1, v_2, \cdots v_n\}$을 $\{f_1, f_2, \cdots f_n\}$ 특징 집합에 클래스 레벨 c와 함께 배정함으로써 기술된다. 만약에 함수 F가 학습되어져야 되면, $c = F(v_1, v_2, \cdots v_n)$이다. 특징 선택은 분류 성능을 최대한으로 하면서 주어

진 원래의 n 특징들에서 m의 유용한 특징들을 선택하는 것이다.

특징 선택 방법들은 평가하기 위해서 무엇을 사용하느냐에 따라 필터와 래퍼 접근 방법으로 나눌 수가 있다. 필터 접근 방법은 분류 과정 전에 인스턴스들 사이에 거리의 측정을 바탕으로 특징들의 부분 집합을 선택하는 것이다. 래퍼 접근 방법은 분류 과정 중에 분류의 결과를 바탕으로 특징의 부분집합을 선택한다.

통계[9], 지리학[10], 기계학습 등을 포함한 다양한 방법들을 통해 연구자들이 시도했던 특징 선택 방법들은 상당히 많이 존재한다.

통계학 방법에서는 전방과 후방 stepwise multiple regression(SMR) 이 특징들을 선택하기 위해 사용된다. 전방 방법은 후방 방법에 비해서 복잡도가 덜하기 때문에 전방 방법이 주로 사용된다.

지리학 방법에서는 탐색 공간에서 인스턴스의 위치들이 결정트리를 위한 특징들을 선택하기 위해서 IDG 알고리즘에 보내진다. 다른 클래스로부터 경계 instance를 분리하는 규칙들은 보상을 받고, 같은 클래스로부터 경계 instance를 분리하는 규칙들은 벌칙을 받는다.

기계학습 방법에서는 sequential forward search (SFS), sequential backward search(SBS)[11] 그리고 변형 방법[12]들이 사용되어졌다. SFS는 빈 집합으로부터 시작해서 집합에 지역적으로(locally) 최고의 특징을 합한다. SBS는 전체 집합에서 시작해서 집합에서 지역적으로 가장 나쁜 특징을 제거한다. 교사 훈련 알고리즘을 가지고 훈련시키는 뉴럴 네트워크 방법들은 오용 침입 탐지에 적용이 되고[13], 비교사 훈련 알고리즘을 갖고 훈련시키는 뉴럴 네트워크는 비정상 침입 탐지에 적용이 된다[14]. 퍼지 집합 이론을 활용해서 FuzzyARTMAP이 적용되고[15], 러프 집합 이론을 가지고 PRESET은 이진 특징들을 선택하기 위해서 특징 집합들의 종속성을 결정한다[16]. GA는 특징 집합을 집합에서 특징들의 존재 유무를 표시하는 1과 0의 비트 스트링으로 코딩해서 특징 선택을 한다[17].

2.2 유전자 알고리즘

GA는 유전적 계승과 다윈적 생존 경쟁이라는 자연 현상을 모델링한 확률적인 탐색방법으로 유전검색이 불가능할 정도로 큰 후보해 공간을 갖는 최적화문제에 적용할 수 있다. 즉, 해가 될 가능성이 있는 개체집단을 유지함으로써 여러 방향의 탐색을 실행하고 이들 방향간의 정보형성과 교환을 행한다. 개체집단은 진화과정을 모방하는데, 각 세대에서 비교적 우량한 해들이 재생산되고, 반면에 비교적 불량한 해들은 소멸된다. 또한 다른 해들간의 차이를 구별하기위해 환경의 역할을 수행하는 목적함수를 사용한다. 이러한 유전자 알고리즘은 특정한 문제에 대해 다섯 가지의 요소를 가져야만 한다. 유전자적 표현방법, 초기 개체집단을 만들어 내는 방법, 목적함수, 유전 연산자, 그리고 여러 가지 매개변수의 값이다.

어떤 개체집단을 초기화하기 위해서는 단순히 개체집단의 염색체를 비트단위로 임의로 설정할 수 있다. 혹은 가능한 최적값들의 분포에 관한 지식을 가지고 있다면 초기의 해집합을 배열하는 데 그 정보를 이용할 수 있다.

알고리즘의 나머지 부분은 각 세대에서 각각의 염색체를 평가하고, 적합도값에 기초한 확률분포에 의하여 새로운 개체집단을 선택하며, 돌연변이와 교배연산자에 의하여 새로운 개체집단의 염색체들을 변화시킨다. 여러 세대 후에 더 이상의 개선이 없으면, 그 세대의 가장 좋은 염색체가 최적해를 나타낸다. 선택과정에서는 적합도에 비례해서 가장 좋은 염색체는 더 많이 복제되고, 보통 염색체는 비슷하게 남아 있으며, 최악의 염색체는 소멸된다. 교배연산자는 교배연산확률을 토대로 두 개의 염색체에 적용되서 새로운 두 개의 자손을 생산하며, 마지막으로 돌연변이 연산자가 돌연변이 확률에 의해 비트별로 적용된다. 이러한 선택,

교배, 그리고 돌연변이를 한 후에 새로운 개체집단은 다음 평가를 받는다. 유전자 알고리즘의 기본 동작은 (그림 1)과 같다.

```
Simple Genetic Algorithms()
{
  initialize population;
  evaluate population;
  while termination criterion not
   satisfied
  {
    perform selection for next
    population;
    perform crossover and mutation;
    evaluate population;
   }
}
```

(그림 1) 유전자 알고리즘의 기본 동작

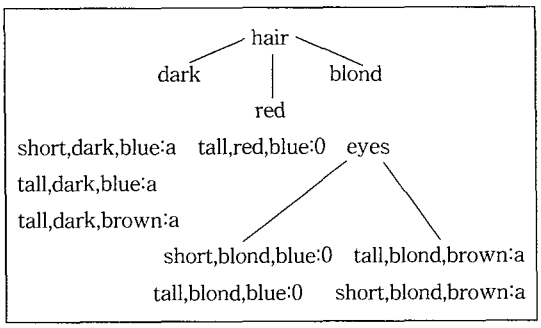
2.3 결정 트리(Decision Tree)

결정 트리는 분류화 작업과 예측을 하는데 있어서 강력하고 가장 많이 사용하는 도구이다. 트리를 기반으로 한 방법인 결정트리나 뉴럴 네트워크의 경우는 규칙(rules)으로 표현이 되는데, 이런 규칙은 읽을 수 있도록 영어로 표현이 된다. 따라서 접하는 사람들은 쉽게 이해할 수 있거나 혹은 데이터베이스에 접근하는 언어인 SQL로 표현이 되는데, 이런 레코드들은 특별한 카테고리를 회수하는 것에는 상당히 약하다. 이러한 응용에서는 분류화나 예측에서의 유일한 문제는 바로 정확성이다.

결정 트리는 전형적으로 루트, 즉 위에서 아래로 향하는 형태를 가지게 된다. 트리의 루트 노드에 레코드를 집어넣게 되면 어떤 자식 노드로 분기를 할 것인지를 결정을 하게 된다. 이러한 절차를 리프 노드에 도달할 때까지 계속적으로 이루어지게 된다. 또한 각각의 리프 노드로 가는 길은 유일하며 그 길은 레코드를 분류하는 규칙으로 표현이 되게 된다.

결정 트리를 구축하는데 있어서 많은 알고리즘이 존재하는데, 그 중에 가장 일반적인 것은 C4.5이다. C4.5와 같은 계통인 ID3는 잠재적인 분기를 비교하여 information gain이라고 불리는 특징을 사용한다. 이것의 중요한 개념은 특정 상황이나 가능한 결과들의 집합의 크기에 의존적인 결과를 많은 수의 비트로 표현하는 것을 요구된다. 만일 여덟 개의 동일하게 예상되는 클래스가 있다면, 어떤 특정한 것을 식별하기 위해서는 세 개의 비트 혹은 $\log_2(8)$이 소요된다. 반면에, 표현된 클래스가 4개만 있다면, 그들 중 하나를 식별하는 데는 $\log_2(4)$ 혹은 두 개의 비트만이 필요하다. 그래서, 노드에서의 분기는 여덟 개의 클래스의 예와 평균 4개의 클래스로 분기하는데, 각각의 클래스는 한 비트의 information gain을 갖는다. 분기 평가 특징으로써 information gain은 매우 무성한 결정 트리로 향하는 높은 성향을 가지고 있다.

2 단계의 ID3 결정트리는 (그림 2)와 같다.



(그림 2) 2단계의 ID3 결정트리

3. GA를 이용한 특징선택

특징 부분집합 선택은 모든 가능한 부분집합에서 분류 성능이 최대화하는 특징들의 부분집합들을 선택하는 과정이다. 넓은 범위의 문제에서 탐색해야 할 탐색 공간은 매우 클 수가 있다. 또한,

특징 선택은 여러 조건과 제약의 최적화 문제이다. 이러한 점이 GA가 특징 선택 문제에 적용하기에 적당하다.

Sklansky등은 전형적인 방법과 비교해서 GA의 우수성에 대한 결과를 제공하였다. 그들은 또한 GA가 NP-hard 문제들을 푸는데 중요한 대안임을 보여주었다. GA 기반의 방법은 Greedy 같은 탐색 방법과 비교되었다.

GA를 활용한 여러 방법들이 제안되었다. 그 중 대부분은 분류 알고리즘의 에러를 최소화해야 할 적응 함수로 사용했으며, 래퍼 접근 방법을 사용하였다. AQ15, ID3/ C4.5, 그리고 K-nearest neighbor 분류 알고리즘 같은 여러 분류기들이 에러 비율을 평가하기 위해서 사용되어졌다. 뉴럴 네트워크과의 결합이 또한 패턴 분류와 IDS를 위한 시스템 구조를 설계하는데 제안되었다[19-23].

특징들의 부분집합을 탐색하기 위해서 각 개체는 통상 n 비트 스트링으로 코드될 수 있는데, 여기서 0 값은 특징 집합에서 제외 된 특징을 1 값은 특징 집합에서 포함되어 있는 것을 나타낸다.

4. 실험 및 고찰

4.1 KDD DATA 분석

실험에서 사용된 데이터는 1999년 "KDD'99 Competition : Knowledge Discovery Contest"에서 제공된 것을 활용하였다. 침입과 정상데이터로 라벨(labels)되어 있는 데이터를 Training Data로 사용하였으며 라벨이 없는 데이터를 Test Data로 사용하였다. 침입 탐지 모듈의 학습을 시키는 작업은 "bad connections" 즉 침입 또는 공격 행위들과 "good connections" 즉 정상사용자들의 행위를 구분 가능한 예측모델(즉 분류기 또는 분류자)을 설계하는 것이다.

1998년 DARPA 침입탐지 개발 프로그램은 MIT Lincoln Labs에서 준비되었고 관리되어져 왔다. 여기서 제공되어진 데이터는 군사 네트워크 환경에서 실험되어진 방대하고 다양한 침입들을 포함하고 있는 표준 감사 데이터집합(data set)들이다. 이후 1999년 KDD Intrusion Detection contest는 바로 이 데이터집합을 활용하여 진행되었다. Lincoln Labs에서 미공군의 LAN에서 9주간의 raw TCP dump data를 얻기 위해 실험 환경을 조성하였다.

Raw Training Data는 7주간의 네트워크 트래픽에서 압축된 Binary TCP dump data 약 4기가 바이트를 사용하였다. 이 데이터는 약 5백만 connection records를 포함하고 있으며 유사하게 2주간의 Test Data는 약 2백만 connection records를 포함하고 있다.

Connection은 잘 정의된 일정한 시간동안 그 처음과 끝이 TCP packets의 연속으로 구성되어 있다. 이것은 신뢰할 만한 프로토콜을 통하여, Packet들의 출발지 IP와 목적지 IP까지의 packets, 그리고 Data의 그 flows를 포함한다. 각각의 connection은 label들이 표시되어 있는데, 정상사용자 인지 비정상사용자(네트워크에 대한 공격, 정확한 공격유형 중 하나)를 표시하는 라벨이다. 각 connection record는 대략 100 bytes로 표현된다. 이 데이터에서 주요한 4개의 공격들의 유형은 다음과 같이 4가지로 분류된다.

- DOS : denial-of-service, 예) syn flood 등등
- R2L : unauthorized access from a remote machine, 예) guessing password 등등
- U2R : unauthorized access to local superuser (root) privileges, 예) various, buffer overflow 등등
- probing : surveillance and other probing, 예) port scanning 등등

여기서 중요하게 생각해야 할 것은, Test data는 Training data와 같이 동일한-공격 유형들의 확률적 분포를 나타내지는 않으며, Training data에는 없는 상세한 공격 유형들을 포함한다. 이러한 것들이 본 실험을 더욱 실제적으로 유용하게 만들어준다. 이러한 Training 데이터집합은 24개의 Training Attack Types을 포함 하고 있으며, 테스트 데이터에는 14개의 공격 유형(types)을 더 포함하고 있다.

Training Attack Types은 다음과 같다.

〈표 1〉 학습 공격 유형

Table 1. Training Attack Types

유 형	Attacks
DoS	back, land, neptune, pod, smurf, teardrop
PROBING	ipsweep, nmap, portsweep, satan
R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster
U2R	loadmodule, buffer_overflow, perl, rootkit

위 공격유형은 탐지 모델에서 각각 탐지하게 된다. 먼저, Dos와 PROBING 공격들은 "Time based traffic model"에서, Slow PROBING 공격은 "host based traffic model"에서, R2L과 U2R은 "content model"에서 각각 생성된 규칙을 사용하여 침입을 탐지 한다.

〈표 2〉 도출된 속성

Table 2. Derived Features

same host features	connections 만을 살펴본 것이다. 2초 동안, 현재 connecton과 동일한 목적지 host를 가지고 있는지, service 등등과 같은 프로토콜 행동들과 관련된 통계량을 계산한 것이다.
same service features	2초간 현재 Connection에서 같은 service를 가지고 있는 connection을 살펴 본 것이다.

공격자들의 connections으로 부터 정상적인 사용자들의 connections을 구별하기에 도움을 주는 higher-level features를 정의하였다. Derived Features의 몇몇 카테고리(categories)가 다음과 같다.

"same host"와 "same service" features 모두 connection records의 time-based traffic features라고 부른다.

호스트(또는 ports)를 스캔하는 PROBING 공격은 매우 오랜 시간을 필요로 한다(2초 이상). 그러므로 connection records는 목적지 호스트로 정렬이 되었으며, features는 time window 대신 동일한 호스트에 100개의 connection windows를 사용하여 구성되었다. 이것은 "host-based traffic features"라 불리는 것들의 집합을 산출해 낸다.

대부분의 DOS와 PROBING 공격은 연속적인 패턴들이 발생하며 매우 짧은 시간동안 여러 호스트(들)에 많은 connections를 수반하지만, R2L과 U2L 공격들의 records에는 연속적인 패턴이 발생하지 않으며 R2L과 U2L 공격들은 packets의 데이터 부분에 포함되어 있으며 정상적인 사용자는 오직 하나의 connection만을 가지고 있다.

packets의 구조화되지 않은 데이터 부분들을 자동적으로 탐색하는 것에 대한 유용한 알고리즘을 연구하는 것은 이미 잘 알려진 연구 분야이다. Stolfo et al.은 데이터 일부에서 로그인을 시도했을 때 실패한 횟수와 같은 행위를 한 의심이 가는 사용자를 찾는 features를 추가하는 기법(또는 지식 분야)를 사용하였다. 그들 feartures를 "content" feartures라 한다.

Features는 각각 KDD Data에 각각의 표현형대로 나열되어 있으며 총 41개의 Features가 있다. Training Data에는 오른쪽 마지막에 라벨이 있다.

4.2 실험 결과 및 분석

이 실험에서 훈련 샘플의 40%로 C4.5 분류기를 구성하는데 사용하였으며 훈련 샘플의 60%는 구

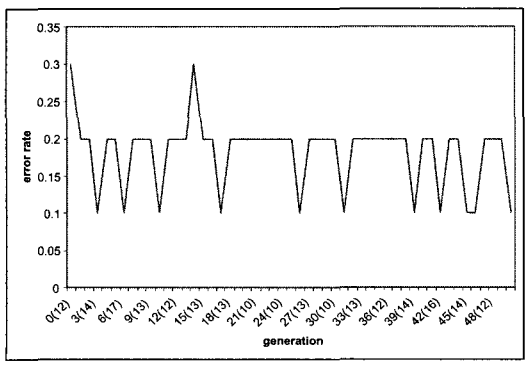
축된 분류기에 의한 분류를 하는 방법으로 에러 비율을 측정한다. 그리고 결과는 알려지지 않은 데이터 샘플에 의해 테스트되어 진다. 즉, KDD 데이터의 훈련 데이터는 494,021 레코드를 가지고 있다. 그 중 197,608 레코드가 이 방법을 훈련하기 위해 사용되며, 296,413 레코드가 이 방법을 테스트하기 위해 사용된다. 방법이 구축 된 후, KDD 데이터의 테스트 데이터는 성능을 평가하기 위해서 사용되어진다.

GA에서는 집단 크기는 $2 \times n(82)$이고 세대의 최대 크기는 50이다(여기서 n은 특징의 수 41). 또한 교체확율은 0.8이고 돌연변이 확률은 0.15이다.

GA에서 초기 집단은 랜덤 방법으로 구축하였으며, 적응함수로서 C4.5 결정트리의 에러 비율을 사용했다. 결과는 (그림 3)에 나타난다. 그림은 각 집단에서 가장 작은 에러 비율을 표현한다. 여러 염색체에서 같은 에러 비율이 나타 날 경우에는 가장 작은 특징 부분집합의 염색체를 선택한다. 각 세대에서 괄호안의 수는 특징의 수이다. 비록 여러 세대에서 0.1의 에러 비율이 발견되었지만, 부분집합의 특징의 수는 10이 넘었다.

보이지 않은 데이터에서 GA의 가장 좋은 결과는 26세대에서 나타났고 염색체는 다음과 같다.

00100111000011000000000100000010001110000



(그림 3) GA를 활용한 특징 선택

부분집합의 11개 특징(3, 6, 7, 8, 13, 14, 24, 31, 35 그리고 37번째 특징)이 테스트 데이터를 분류하기 위해 선택되어졌으며 에러 비율은 0.1이다. 보이지 않은 데이터에서 실험을 수행하였다.

훈련 데이터로 원래의 데이터 확장 파일을 사용하였고 테스트 데이터로 보이지 않은 데이터를 사용하였다. 비록 특징들의 수가 11이지만 SGA의 에러 비율은 8.8%이다.

5. 결 론

침입탐지시스템은 최적 조건을 바탕으로 침입을 탐지하기 위해서 가치있는 특징을 선택하는 문제에 직면해 있다. 과거 40여 년간 집중적인 연구가 진행했지만, 특징 선택 문제는 효율적인 시스템을 구축하는데 큰 장애가 되고 있다.

이 논문에서는 특징 선택 문제에 대해 GA의 적응 함수를 사용해서 수행하였으며, 적응함수로는 결정트리의 에러 비율을 활용하였다. 실험을 통해 적은 수의 특징으로 나은 결과를 발견했다.

앞으로 최적의 방법과 조건에 맞는 분류 방법을 연구할 것이다. 특징들의 연관 관계와 다른 특징에 비해 중요한 특징들에 대한 연구도 진행한다. 더불어, 제안하는 방법은 이론적 배경과 실질적인 매우 큰 범위의 데이터에 적용되어져야 한다.

참 고 문 헌

[1] Denning, Dorothy E., An Intrusion Detection Model, IEEE Transaction on Software Engineering, Feb 1987.

[2] Langley, P., Selection of Relevant Features in Machine Learning, In Proc. of the AAAI Fall Symposium on Relevance, New Orleans, LA : AAAI Press, 1994.

[3] W. F. Punch, E. D. Goodman, M.Pei, L. Shun, P. Hovl, and R. Enbody, Further Research on Feature Selection and Classification Using Genetic Algorithms, ICGA93, pp. 557-564, 1993.

[4] Jihoon Yang and V. Honavar, Feature Subset Selection Using a Genetic Algorithm, IEEE Intelligent Systems, pp. 44-49, 1998.

[5] H. Vafaie and K. De Jong, Genetic Algorithms as a Tool for Feature Selection in Machine Learning, Proceeding of the 4thInternational Conference on Tools with Artificial Intelligence, Arlington, VA, November, 1992.

[6] M. J. Martin-Bautista and M. A. Vila, A Survey of Genetic Feature Selection in Mining Issues, Proc.1999 Congree on Evolutionary Computation(CEC '99), pp. 1314-1321, July 1999.

[7] Il-Seok Oh, Jin-Seon Lee and Byung-Ro Moon, Hybrid Genetic Algorithms for Feature Selection, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 26, No. 11, November 2004.

[8] M. Kudo and J. Sklansky, Comparisonof algorithms that select features for pattern classifiers, Pattern Recognition, 33, pp. 25-41, 2000.

[9] Kittler, J., Mathematical Methods of Feature Selection in Pattern Recognition, International Journal of Man-Machine Studies, 7, pp. 609-637, 1975.

[10] Elomaa, T. and E. Ukkonen, A Geometric Approach to Feature Selection, In Proceedings of the European Conference on Machine Learning, pp. 351-354, 1994.

[11] Devijver, P. A. and J. Kittler, Pattern Recognition : A Statistical Approach, Prentice -Hall, 1982.

[12] Pudil, P., F. J. Ferri, J. Novovicova and J. Kittler, Floating Search Methods for Feature Selection with Nonmonotonic Criterion functions, IEEE 12th International Conference on Pattern Recognition, Vol. II, pp. 279-283, 1994.

[13] R. P. Lippmann and R. K. Cunningham, Improving intrusion detection performance using keyword selection and neural networks, Computer Networks, Vol. 34, No. 4, pp. 597-603, 2000.

[14] A. J. Hoglund, K. Hatonen, an A. S. Sorvari, A computer host-based user anomaly detection system using the self-organizing map, in Proc. of the IEEE-INNS-ENNS Int. Joint Conf. on Neural etworks(IJCNN2000), Como, Vol. 5, pp. 411-416, 2000.

[15] J. Cannady and R. C. Garcia, The application of fuzzy ARTMAP in the detection of computer network attacks, in Artificial Neural Networks-ICANN 2001.

[16] Modrzejewski, M, Feature Selection Using Rough Set Theory, European Conference on Machine Learning, pp. 213-226, 1993.

[17] Siedlecki, W. and J. Sklansky, A Note on Genetic Algorithms for Large-scale Feature Selection, Pattern Recognition Letters, 10, 5, pp. 335-347, 1989.

[18] J. H. Holland, Adaptation in Natural and Artificial Systems, Univ. of Michigan Press, Ann Arbor, Mich., 1975.

[19] W. Siedlecki and J. Sklansky, On Automatic Feature Selection, International Journal of Pattern Recognition and Artificial Intelligence 2(2), pp. 197-220, Singapore : World Scientific Publishing Company, 1988.

[20] Vafaie H. and Imam, F., Feature Selection Methods : Genetic Algorithms vs. Greedy -like Search, In Proceedings of the 1994

International Fuzzy Systems and Intelligent Control Conference, Louisville, KY, 1994.

[21] Vafaie. H. and De Jong, K., Genetic Algorithms as a Tool for Restructuring Feature Space Representations, In roceeding of the Seventh International Conference on Tools with Artificial Intelligence, Herndon, VA, 1995.

[22] Cherkauer, K. J. and Shavlik, J. W., Growing Simpler Decision Trees to Facilitate Knowledge Discovery, In Proc. of the second International Conference on Knowledge Discovery and Data Mining(KDD - 96), pp. 315-318, Portland, OR : AAAI Press.

[23] Alexander, Timo Horeis, and Bernhard Sick, Feature Selection for Intrusion Detection : An Evolutionary Wrapper Approach, IEEE pp. 1563-1568, 2004.

한 명 묵

1980년 연세대학교 공과대학 (공학사)

1987년 뉴욕공과대학교 컴퓨터 공학과(공학석사)

1997년 오사카시립대학교 정보 공학부(공학박사)

1998년~현재 경원대학교 소프트웨어대학 부교수