RBF 신경망을 이용한 비정상 행위의 탐지 기법

김형태, 김영호, 이금순, 강주미, 원용관
전남대학교 컴퓨터공학과
e-mail:{timeless, ykwon}@grace.chonnam.ac.kr

Anormal Behavior Detection Using RBF Neural Network

H. T. Kim, Y. H. Kim, K. S. Lee, J. M. Kang, Y. Won
Dept of Computer Engineering, Chonnam National University

요약

컴퓨터 시스템 및 네트워크에 대한 침입 공격의 방법 중 이미 알려진 형태의 공격에 대해서는 상대적으로 탐지가 용이하나, 사용자의 비정상행위는 방법의 다양성 때문에 탐지가 매우 어렵다. 그러나, 사용자의 정상적인 행동은 몇자지 소수의 형태로 특정 지어질 수 있다. 본 논문에서는 상대적으로 변화가 적은 정상 행위를 신경망으로 Modeling하여 이를 비정상 행위 탐지에 적용하는 기법을 제안한다. 이를 위하여 입력 영역을 지역화하는 특성을 갖는 RBF(Radial-Basis-Fuction) 신경망에 대한 단일 Class의 학습 방법을 제안하고, 이를 이용한 비정상 행위에 대한 공격의 탐지를 위한 적용 방안을 제시한다. 비정상 행위 탐지에 대한 적용 가능성을 검증하기 위하여 각 사용자의 키보드 입력 유형을 학습하고 이를 이용하여 타인의 ID와 Password를 도용한 경우의 탐지에 적용하였다.

1. 서론

정보호보란, 정보통신망에서 처리되는 정보의 기밀성하고 시스템의 가용성을 보장하는 기술로서 재산권, 시민권 및 국가 경제와 사회 보호에 중요한 부분을 담당하고 있다. 정보보호의 방법에는 다음과 같은 두 가지 방법이 있다.

○ 비정상적인 행위 탐지(Anomaly Detection) : 정상적인 system 사용에 관한 profile과 system 상태를 유지하고 있다가, 이 profile에서 벗어나는 행위를 관찰하고, 각각의 행위에 대한 profile을 생성한다. 생성된 profile들을 주기적을 관찰하여 profile의 비정상적인 행위의 정도(abnormality)를 측정한다.

○오용 침입탐지(Misuse Detection) : system의 알려진 취약점들을 이용하여 공격하는 행위들을 사전에 공격에 대한 특징 정보를 가지고 있다가 탐지하는 방법이다.

본 논문에서는 단일 Class만을 대상으로 하는 Supervised 학습 방법을 이용하여 사용자의 비정상 행위를 탐지하고자 한다. 먼저 입력 영역의 지역화 특성을 갖는 RBF와 MLP을 결합한 새로운 신경망 구조를 소개한다.

본 논문의 구성은 다음과 같이 구성이 되어 있다. 2장에서는 RBF-MLP 관련 이론을 살펴보고, 3장에서는 RBF-MLP 신경 회로망의 구조를 제안하고 단일 class에 대한 학습 규칙을 소개한다. 4장에서는 비정상 행위 탐지에 대한 적용 방안의 가능성을 보이기 위하여 특정 단어의 타자 습관을 이용하여 시스템 사용자에 대한 authetication 적용의 실험과 결과 및 분석을 제시한다. 마지막으로 5장에서는 결론과 관찰된 문제점에 대한 대안을 제시한다.

2. RBF 신경 회로망 모델 및 관련 연구

2.1 RBF 신경회로망

RBF들은 특정 함수들의 class이다. RBF들의 특징은 중심으로부터 거리가 멀어질수록 함수 값들이 단조증가 또는 감소한다는 것이다. 즉, 입력 영역의 지역화 특성을 갖는다. 이런 특징으로 인해 RBF들은 대체로 지역성의 특징이 있다.

RBF Network은 다음에 살펴볼 K-Means algorithm을 사용하여 k개의 Clustering을 구성을 하고, 각각의 Clustering에 대하여 Input Vector Space에 Gaussian function을 사용하여 Input Vector가 각 Cluster에 포함된 정도를 보여준다. RBF의 신경망 모델은 <그림 2>와 같이 구성이 되어 있다. 그림에서 구성된 RBF의 Kernel Function은 식 (2.1)의 Basis Function을 사용한다. [3][4][8]

$$f_m(x^{(q)}; v^{(m)}) = \exp\left[-\frac{\| x^{(q)} - v^{(m)} \|^2}{2\sigma_m^2}\right] \quad (2.1)$$

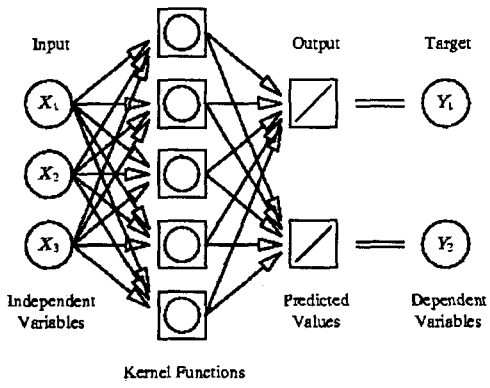


그림 2 Radial Basis Function(RBF) Network

2.2 Error-Backpropagation 학습 (오류 역전파 학습)

Backpropagation algorithm은 출력값과 기대값을 비교해 차이를 줄여나가는 방향으로 가중치를 조정해 나간다. 이 algorithm은 목표값(d)와 실제 출력값(o) 사이의 오차 제곱합(sum of squared errors)으로 정의된 함수(fuction) E의 값을 경사하강 추적법(gradient descent method)에 의해 최소화하는 방향으로 학습한다. [1][4][8]

$$E = \sum_p E_p, \quad (E_p = \frac{1}{2}\sum_j (d_{pj} - o_{pj})^2) \quad (2.2)$$

이 학습 규칙은 실제 출력과 목표 출력간의 평균 제곱 오차(Mean Square Error : MSE)를 최소화시키기 위해서 , 연결 가중치에 대한 오차의 미분계수가 감소하는 방향으로 연결가중치를 계속 변경시키는 반복적인 경사하강(interative descent) 알고리즘이다. 식(2.2)에서 각 Layer의 연결 가중치의 변화를 구하기 위해서는 Chain Rule을 사용한다.

$$\frac{\partial E_p}{\partial w_{ji}^h} = \frac{1}{2}\sum_k \frac{\partial (y_{pk} - o_{pk})^2}{\partial w_{ji}^h}$$

$$= -\sum_k (y_{pk} - o_{pk}) \frac{\partial o_{pk}}{\partial net_{pk}^o} \frac{\partial net_{pk}^o}{\partial i_{pj}}$$

$$\times \frac{\partial i_{pj}}{\partial net_{pj}^h} \frac{\partial net_{pj}^h}{\partial w_{ji}^h} \quad (2.3)$$

따라서, (2.3)에서 p번째 Pattern의 Weight 변화는 공식(2.4)과 같이 표현을 할 수가 있다.

$$\Delta_p w_{ji}^h = \eta f_j^{h'}(net_{pj}^h) x_i \sum_k \delta_{pk}^o w_{pk}^o \quad (2.4)$$

2.3 K-Means

C-means clustering 이라고도 알려진 K-means clustering은 영상과 음성 Data 압축을 포함하여 radial basis function network를 사용한 시스템 Modeling을 위한 데이터 전(前)처리와 異種신경망 구조 안에서 분리 작업을 한다.

K-means 알고리즘은 벡터 x_j(j=1,2,3,....,n) n 개의 집합을 Group G_i(i=1,2,3,...,c) C개로 분할하며, dissimilarity (distance) measure의 cost(objection) function이 최소화되는 각 Group안의 cluster 중심을 찾는다. k-means 알고리즘의 수행은 cluster center의 초기 위치에 따라 달라진다. 그러므로 좋은 초기 cluster center를 찾기 위해 front-end방법을 사용하거나 각기 다른 초기 k-means 알고리즘 의 집합을 가지고, 여러 번 알고리즘을 구동하여 본다.

3. RBF-MLP 신경 회로망

3.1 신경 회로망 구조

본 논문에서 설계할 RBF-MLP 신경회로망은 RBF와 MLP가 Cascade 형태로 연결된 구조의 형태를 가지고 있다. 다음 <그림 3>는 RBF-MLP 신경 회로망 구조를 보여주고 있다.

Input Space는 사용자의 Pattern Class를 앞 절에서 살펴보았던 K-Means Clustering을 사용하여 k개의 Cluster로 구분하는 전처리 단계와 RBF의 Kernel Function이 함께 존재한다. [3][7][8]

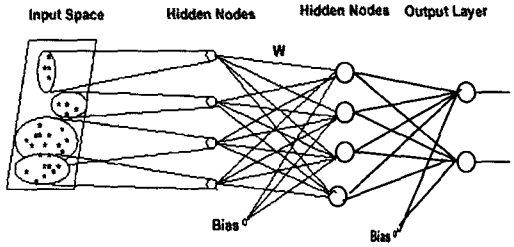


그림 3 RBF-MLP 신경 회로망 구조

Input Space에서 첫 번째 Hidden Node로 전달되

는 부분은 RBF가 입력 패턴에 대하여 MLP에 대한 특징 추출의 단계로 RBF의 출력을 MLP는 입력으로 받아들인다. 또한 각 layer의 연결가중치(weight)를 error backpropagation algorithm으로 학습을 시킨다.

3.2 단일 Class Data의 학습

신경회로망이 형태 분류에 적용되는 경우 다중 클래스에 대한 분류를 수행하도록 학습을 시킨다. 정상 행위와 비정상 행위의 판정은 Tow-Class 분류 문제이나 비정상 행위를 충분히 표현 할 수 있는 데이터의 확보가 곤란하다. 또한, 정상 행위 데이터만을 가지고 학습을 실행할 경우 신경 회로망 시스템은 비정상 행위를 구분할 수 있는 능력이 없다. 따라서, 정상 행위 및 비정상 행위를 충분히 표현할 수 있는 학습 데이터의 확보가 가장 중요하다.

만약, 탐지 시스템에 대하여 비정상 행위에 해당하는 입력 데이터에 대하여 낮은 출력 값을 만들어 내도록 정상 행위에 대한 학습이 이뤄질수 있다면 오직 정상행위 데이터 만으로도 비정상행위를 탐지할 수 있는 시스템 학습이 가능해질 것이다. 이 경우, 정상 행위의 데이터는 비정상 행위보다는 상대적으로 확보가 쉽고 또한 변화도가 적어 학습이 쉽게 이루어 질수 있다.

정상 데이터에 대한 단일(Single) Class의 학습은 입력 공간(Input Space)에서 정상 패턴 영역을 지역화하도록 RBF 신경망을 학습 시키는 것이다. 이는 RBF의 기저함수(Basis Function)들이 최소의 Volume으로 입력 데이터의 군집(Cluster) 영역에 위치하도록 학습을하여 달성할 수 있다. 즉, RBF의 기저함수가 Gaussian Function일 경우 평균이 정상 패턴의 Cluster 중앙에 위치하고 분산 값이 최소화 되도록 학습을 수행하는 것이다.

이러한 목적을 달성하기 위하여 식 (3.1)과 같은 목적함수(Objective function)를 정의하고 이를 최소화하도록 학습을 수행토록 한다.

$$E=\frac{1}{2}\left(\sum_j (d_j - o_j)^2 + \sum_i \sigma_i^2\right) \qquad (3.1)$$

3.3 학습 규칙(Learning Rule)

MLP에 대한 일반적인 학습 규칙은 널리 알려져 있다. 이 장에서는 RBF의 파라미터인 중앙값 및 분산 값에 대한 학습을 소개한다. 이것은 RBF가 최소의 Volume으로 입력 데이터의 Cluster 영역에 위치시키기 위하여 학습을 시키는 것이다.

학습 규칙은 RBF 계층을 하나의 은닉층으로 간주하고 Delta(Chain) Rule을 적용하여 구한다. 식 (3.1)의 목적함수(Object function) 즉, TSSE(Total Sum of Squared Error)에서 learning rule이 이루어진다.

다음은 variance의 learning rule의 공식이다.

$$\frac{\partial E}{\partial \sigma_i} = \frac{\partial E}{\partial net_j} \times \frac{\partial net_j}{\partial \sigma_i} \qquad (3.2)$$

$$= G(x) \times \sum_i \frac{\| x_p - c_i \|^2}{\sigma_i^3} \times w_{ji} \qquad (3.3)$$

따라서,

$$\Delta \sigma_i = \eta_1 \times G(x) \times \sum_i \frac{\| x_p - c_i \|^2}{\sigma_i^3} \times w_{ji} \times \delta_j + 2\sum_i \sigma_i \quad (\eta_1 \text{학습율}) \qquad (3.4)$$

Center에 대한 learning rule을 구하면,

$$\frac{\partial E}{\partial c_{ic}} = \frac{\partial E}{\partial net_j} \times \frac{\partial net_j}{\partial c_{ic}} \qquad (3.5)$$

$$= G(x) \times \sum_i \sum_c \frac{(x_{pc} - c_{ic})}{\sigma_i^2} \times w_{ji} \qquad (3.6)$$

따라서,

$$\Delta c_{ic} = \eta_2 \times G(x) \times \sum_i \sum_c \frac{(x_{pc} - c_{ic})}{\sigma_i^2} \times w_{ji} \times \delta_j \qquad (3.7)$$

여기서, x_{pc} 는 cluster에 해당되는 pattern들의 좌표 포인트이고, c_{ic} 는 center의 각 좌표 포인터 들이다.

Weight에 대한 learning rule을 구하면,

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j} \times \frac{\partial net_j}{\partial w_{ji}} \qquad (3.8)$$

$$\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial \sum_k w_{ji}\, o_i}{\partial w_{ji}} = o_i \qquad (3.9)$$

따라서,

$$\Delta w_{ji} = \eta_3 \times \delta_j \times o_i \qquad (3.10)$$

(단, o_i 가 Gaussian function일 경우는 Input layer와 Hidden layer사이에서만 적용이 된다. 그 이외에는 Sigmoid function이 적용이 된다.)

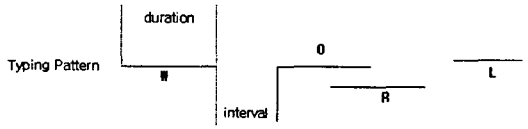
4. 실험 결과 및 분석

4.1 실험 데이터 수집

실험 데이터의 수집은 특정 단어 및 단어군에 대

하여 인가된 사용자의 타이핑 특성을 수집한다. 타이핑 패턴의 특징 추출에 이용한 단어는 "worldwide", "Computer secutity" 및 "Database Management Systems"이다. 사용자의 Keystroke 특징은 시간의 값으로만 구성된 Time Vector로써, 문자의 해당키를 누르고 있는 "키 지속 시간(keystroke duration time)"과 현 키의 입력이 유효하지 않은 시간부터 다음 키의 시간까지의 차이 값인 "키 사이 시간(key interval time)"로 정의된다. 위의 <그림 4>는 keystroke duration time과 key interval time을 사용하여 Time Vector를 생성하였다. [2][4]

본 논문에서는 key interval time에 대해서만 특징 추출을 한다면 N-1차의 시간 벡터를 구성을 한다. 이 시간 벡터에서 음의 시간 값은 키 사이 시간에만 나타나며, 그 의미는 현재 눌려진 키가 떨어지기 전에 뒤따르는 키가 눌려졌음을 의미한다. [2][5]



Timing Vector = [60,60,90,-30,10,30,60]

그림 4 "WORL"의 타자패턴

4.2 실험 결과 및 분석

사전에 선정한 단어에 대하여 인가된 사용자가 50회 반복적으로 타이핑을 수행하여 패턴의 특징을 추출하여 이를 학습에 적용한다.

UNIX 계열에서 비 인가자가 사용할 경우 치명적 문제를 발생하는 명령어들(예: rm, mv, shutdown, reboot 등)의 실행을 요청할 경우 2차적 인증을 수행한다. 2차적 인증은 사전에 특징 추출에 이용된 단어 중 임의로 하나를 제시하고 이를 타이핑하도록 하여 제안한 시스템으로 판별한다. 인가된 사용자와 5명의 비인가 사용자로 하여금 각각 재 인증을 요구하는 명령어를 15회씩 수행토록 하여 시스템의 성능을 평가하였다.

<표 1>을 살펴보면 비 인가자를 인지하는데는 성공적이였다. 그러나, 인가자를 비 인가자를 분류하는 경우가 있는데 -오분류 율 26.6% - 이는 50회를 연속적으로 타이핑할 경우에는 패턴이 균일화되지만, 일정시간 시간이 지나서 동일한 단어를 타이핑할 지라도 패턴의 균일성이 낮아지는데 그 원인이 있다고 분석할 수 있다.

	인가자	비인가자
인가자	11	4
비인가자	0	75

표 1 성능 평가 실험 결과

5. 결론

본 논문에서는 RBF의 지역화 특성과 MLP 비선형 분류 능력의 장점을 이용하여 단일 클래스 분류 학습 시스템을 제안하였다. 제안된 시스템을 활용하여 시스템 사용자의 재 인증에 적용, 비인가자에 대한 비정상 행위를 탐지하는데 성공적인 적용 가능성을 보였다. 하지만, 인가자를 비 인가자로 분류하는 False negative의 비율(26.6%)이 높았다. 이를 개선하기 위해서는 인가자로부터 패턴 데이터 추출 방법의 개선과 많은 데이터의 수집이 요구된다.

참고문헌

[1] 김대수."신경망 이론과 응용" 하이테크정보 1999.

[2] 유진승 "암호타자 패턴의 특징을 기반으로 한 가지 연상 다층퍼셉트론 신경망을 이용한 사용자 인증" 1999

[3] Carl G.Looney "Radial Basis Functional Link Nets as Learing Fuzzy Systems"

[4] M.S Obaidat and Balqies Sadoun "Verification of Computer Users Using Keystroke Dynamics" IEEE Transactions and Systems, Man ad Cybernetics -Part B: Vol 27, No.2 April 1997

[5] John A. Robinson, Vicky M. Liang, J. A. Michael Chambers, and Christine L.MacKenzie "Computer User Verification Using Login String Keystroke Dynamics" IEEE Transactions and Systems, Vol 28, No.2 1998

[6] M.S Obaidat, Senior Member, IEEE, and D.T.-Macchairolo"A Multilayer Neural Network system for Computer Access Security" IEEE Transactions and Systems, Vol 24, No.5 May 1994

[7] Christopher M. Bishop "Neural Network for Pattern Recognition" OXFORD UNIVERSITY PRESS 1995

[8] DON.R HUSH and Bill G. Horne "Progress in Supervised Neural Networks" IEEE Signal Processing Magazine, January 1993