

입국 심사 시 신경망 기법을 이용한 잠재적 테러리스트의 식별

김성혁*, 조경수, 김응모
*성균관대학교 정보통신공학부
e-mail : na062882@gmail.com

Identifying Potential Terrorists by using the Neural Network

Seong Hyeok Kim*, Kyung Soo Cho, Ung-mo Kim
School of Information and Communication Engineering,
Sungkyunkwan University

요 약

2001년 9.11 사태 이후, 미국과 우방국들은 본격적으로 테러와의 전쟁을 시작하게 되었다. 국제화 시대의 흐름에 따라 국내의 외국인의 영향력이 커져가고 있고, 2011년 11월 11일 서울에서 개최되는 G20 세계정상회의를 앞둔 시점에서 테러에 대한 사전 대응이 중요해졌다. 이를 위해 민·관·군이 협력하여 다양한 정책과 활동을 벌이고 있는데, 데이터 마이닝의 기법 중 신경망 기법을 이용하여 효율적이고 기계적으로 잠재적 테러리스트를 식별하는 방법을 제안한다. 신경망 기법은 인간의 두뇌를 모델로 등장하였고, 입력 자료를 통해 학습한다는 것이 특징이다. 이 신경망을 여러 기관에서 발표한 테러리스트들의 정보를 입력 자료로 변환하여 학습시킨 뒤, 검증을 거쳐 실전에 적용하는 것이 본 논문에서 제안하는 신경망 기법 적용의 목표이다. 개인정보보호에 관한 법률 및 방법론이 보완된다면 실전에서 유용하게 사용될 수 있을 것이다.

1. 서론

2001년 9월 11일, 미국 뉴욕의 세계무역센터 쌍둥이 빌딩과 워싱턴 D.C의 국방부 펜타곤이 민간기를 납치한 이슬람 테러단체에 의해 자살테러를 당했다. 이로 인해 90여 개국 2,800~3,000명의 인명피해와 함께 세계무역센터 건물 가치 11억 달러, 테러 응징을 위한 긴급 지출안 400억 달러, 재난극복 연방 원조액 111억 달러 등의 재산상 피해와 그 외의 여러 가지 국가적, 경제적 피해를 입게 되었다. 이에 미국과 그의 우방국들은 본격적으로 테러와의 전쟁을 시작하게 되었고, 2010년 7월 1일 우리나라는 아프가니스탄 파병을 시행하게 되었다.

국내 입국자 중 외국인은 2007년 6,425,257명, 2008년 6,823,812명, 2009년 7,833,998명으로 매년 증가 추세에 있고, 2010년 1분기 현재 총 체류외국인은 1,180,598명으로 자국 내 외국인의 영향력은 점점 커져가고 있다. 이는 세계화의 완성도를 보여주는 것이기도 하지만, 국내 테러 발생의 높은 가능성을 보여주는 것이기도 하다.[1]

2010년 11월 11일 서울에서 개최되는 G20 정상회의를 비롯하여 앞으로 열릴 많은 국제적 행사에서 발생할 수 있는 테러의 사전 방지 및 대응책에 대한 중요성이 강조되고 있다. 이에 대하여 정부에서는 민·관·군 합동 모의 훈련과 공항에의 최신 검색장비 설치 등의 움직임을 보이고 있다.

본 논문에서는 데이터 마이닝의 기법 중 하나인 신경

망 기법을 소개하고, 이 기법을 공항의 입국 심사에 적용하여, 빠르고 효율적이며 기계적으로 테러리스트 의심 인물을 찾아내는 방법을 제안하려고 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 연구에 대해 알아본 뒤, 3장에서는 본 논문에서 제시하는 신경망 기법을 입국 심사에 어떻게 적용하는지에 대해 설명한다. 4장에서는 본 논문의 전체적인 설명 및 제안에 대한 결론과 향후 연구 방향에 대해 제시하며 결론을 맺는다.

2. 관련연구

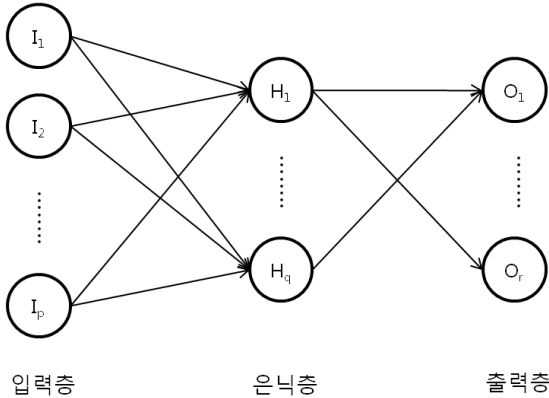
2.1 데이터마이닝

데이터마이닝이란 방대한 양의 데이터 속에서 쉽게 드러나지 않는 유용한 정보를 찾아내는 과정이라고 말할 수 있다. 대용량의 자료를 사용하고 컴퓨터 중심적 기법을 활용하며 통계학, 패턴인식, 뉴로 컴퓨팅, 인공지능, KDD 등의 여러 분야를 망라하고, 이론적 원리보다는 경험에 기초하였다는 특징이 있다.[2]

2.2 신경망 기법

신경망 기법은 인간의 두뇌를 모델로 하여 등장하였다. 사람들이 경험을 통해 학습하는 것과 같이 신경망은 예제를 통하여 학습하는 능력이 있다.

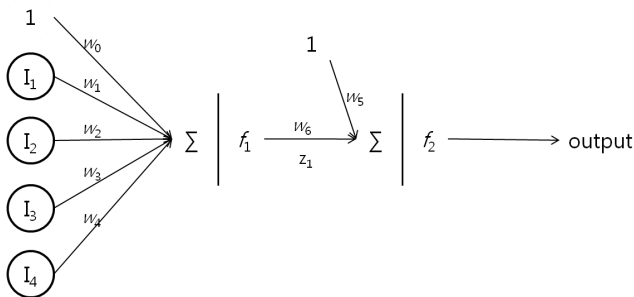
2.2.1 신경망 구조



(그림 1) MLP 신경망 구조

신경망에는 여러 가지 모형이 있고, 가장 널리 사용되는 모형은 MLP(Multilayer Perception) 신경망이다. (그림 1)은 p개의 노드로 구성된 입력층, q개의 노드로 구성된 은닉층, r개의 노드로 구성된 출력층으로 이루어져 있다. 입력층은 각 입력 값에 대응되는 노드들로 구성되고, 은닉층은 입력층으로부터 전달되는 변수 값들의 선형결합을 비선형함수로 처리하여 출력층으로 전달하며, 출력층은 목표 변수에 대응하는 노드들로 구성된다.[2]

2.2.2 신경망의 출력 과정



(그림 2) 결합함수와 활성화함수

(그림 2)에서의 최종 결과 output이 계산되는 과정은 다음과 같다.

$$output = f_1(w_5 + w_6 z_1) \quad \text{식 (1)}$$

$$z_1 = f_2(w_0 + w_1 I_1 + w_2 I_2 + w_3 I_3 + w_4 I_4)$$

식 (1)에서 \$z_1\$은 입력변수들의 선형결합을 활성화함수로 변환한다. 대부분의 신경망에서는 이와 같은 선형함수를 사용한다. (그림 2)에서의 추가적인 노드는 외부 입력 뿐만 아니라 바이어스(bias)도 작용하므로 이를 고려해서 추가되어 신경망이 패턴을 더 잘 이해하도록 해준다. 활성화함수 \$f_1, f_2\$는 일반적으로 시그모이드 함수(sigmoid function)를 사용한다.

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad \text{식 (2)}, \quad f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad \text{식 (3)}$$

식 (2)는 단극성 시그모이드 함수이다. \$\lambda\$는 경사도이고 함수 값의 범위는 \$0 < f(x) < 1\$이다. 식 (3)은 양극성 시그모이드 함수이고 함수 값의 범위는 \$-1 < f(x) < 1\$이다.[2]

2.2.3 입력 자료의 변환

입력 변수를 변환시켜 [0,1] 사이의 값을 가지도록 하는 것이 중요하다. 어떤 구간 사이의 값을 모두 가질 수 있는 연속형 변수는 [0,1] 사이로 변환하기 쉽다. 더 좋은 결과를 얻기 위해서는 자료의 범위를 좀 더 넓게 고려하고, 훈련이 시작된 후 범위를 벗어난 자료의 입력에 대해서는 사용하지 않거나 0 또는 1로 간주한다.[2]

2.2.4 Back Propagation Network

Back Propagation(BP)은 신경망을 학습시키기 위해 현재 가장 널리 사용되는 탐색 기법이다. 초기 단계에서 BP는 자신의 입력부에 무작위의 가중치를 할당한다. 이 신경망을 학습시키고자 할 때에는 학습용 입력 데이터 집합을 적용한다. 각 학습 예제 데이터에 대해 실제 출력과 정확한 값에 해당하는 기대 출력 값을 비교하고, 여기에 차이가 있을 때에는 각 노드와 네트워크의 스냅스에 대한 가중치를 조정한다. 이러한 과정은 신경망의 구조가 안정화 되면 종료가 되며 분석에 사용할 수 있게 된다.[3][4]

3. 신경망 기법을 이용한 테러리스트의 탐지

3.1 제언 배경

테러는 명백히 테러리스트에 의해 자행된다. 테러리스트는 UN, 미국, 영국에서 발표한 "Terrorist" list[5]의, 그리고 그 외의 외국인들과 출국 경험이 있는 자국민들로 구성된다. 이들에게는 테러 활동을 수행하기 위해 입국 심사가 처음이자 중요한 관문이 될 것이고, 역시나 대 테러 활동을 맡은 정부나 기관에도 테러의 사전 방지 차원에서 중요한 기회가 될 것이다.

입국 심사는 국가안보와 위기관리 차원에서 중요한 활동이기에 시간이 오래 걸리고, 또한 입국 심사관에 의존해야 하는 등의 단점을 가지고 있다. 이것을 좀 더 효율적이고 기계적으로 수행하기 위해서 신경망 기법 적용을 제안한다.

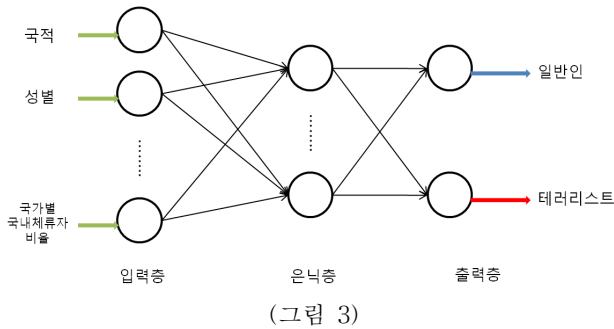
3.2 동작 시나리오

잠재적 테러리스트를 탐지하기 위해서 그동안 발생했던 사건들과 그에 관여한 테러리스트들의 개인정보를 (그림 1)의 입력층 노드의 입력 값으로 신경망을 학습시킨다. 이후, 검증 자료로 테스트를 거치면 실제 적용이 가능해진다. 입국 심사 시, 심사 대상자의 개인정보가 입력되면 잠

재적 테러리스트의 가능성을 수치로 출력하게 된다.

3.3 제안 내용

3.3.1 제안된 신경망 구조



3.3.2 제안된 학습 자료

신경망의 적용에 있어서 가장 중요한 것이 ‘학습’이다. 학습이 제대로 이루어지지 않으면 출력결과를 신뢰할 수 없기 때문이다. 좋은 학습을 위해서는 좋은 입력 데이터를 선정해야 하는데, [0,1] 사이의 값을 가져야 하기 때문에 적절한 변환이 필요하다. 일반적으로

$$\text{변환값} = (\text{실제값} - \text{최솟값}) / (\text{최댓값} - \text{최솟값}) \quad \text{식 (2)}$$

으로 구할 수 있다. 만일 분포가 한쪽으로 치우치는 경우에는 log 변환으로 분포를 균등하게 만들 수 있다.

<표 1> 입력 변수의 종류, 범위, 예시, 변환

변수	변수 값의 범위	예시	변수 값의 변환
국적	4-894	408	0.46
성별	0-1	0	0
나이	0.1-0.9	29	0.29
경유국가	4-894	368	0.41
해당국가 테러발생 비율	0-100	13	0.13
경유국가 테러발생 비율	0-100	35	0.35
국가별 국내 입국자 비율	78-743,273	23,037	0.031
국가별 국내 체류자 비율	4,352 - 563,479	10,098	0.018

<표 1>에서 국적은 ISO 국가 코드로 변수 값의 범위를 만들었다. 예시 408은 대한민국의 국가 코드이고 입력 값의 변환은 식 (2)를 이용하였다. 성별은 남자는 0, 여자는 1로 변환하였고, 나이는 10대에서 90대를 범위로 정하여 변환하였다. 경유국가 역시 국적과 같은 방식으로 변환하였고, 해당국가 테러발생 비율, 경유국가 테러발생 비율은 테러정보종합센터의 ‘월간테러동향’[6] 자료에 기초하여 발생 비율로 변환하였다. 국가별 국내 입국자, 국가별 국내 체류자의 범위는 출입국 외국인정책본부 통계자료[1]에 기초하여 범위를 정하고 식 (2)를 이용하여 변환하였다.

3.3.3 결과 값의 재해석

입력 값을 [0,1]사이의 범위로 변환하였기 때문에 출력되는 결과 값도 [0,1]사이의 범위이다. 그렇기 때문에 그 결과 값을 재해석 하는 것은 중요한 일이다. 본 논문에서 제시한 신경망의 구조는 출력 결과가 두 가지로 나타나게 되어있다. 실제 출력되는 결과 값은 [0,1]사이의 값이기 때문에, 0보다 1쪽에 더 가까운 값이라면 1로 해석할 수 있다. 그러나 만일 결과 값이 0.5에 근접한다면 정확한 판단이 불가능해질 것이다. 이런 경우에 대비해서 이분법적인 판단이 아닌 신뢰도(가능성)를 재해석의 기준으로 사용할 수 있다. 예를 들어, 결과 값이 0.49가 나온다면 0일 신뢰도(가능성)는 51%가 되는 것이다. 반대로 1일 신뢰도(가능성)는 49%가 되는 것이다. 제안된 신경망에 개인 정보를 입력 값으로 적용했을 때 결과 값이 0.49가 나왔다면 일반인(신뢰도 51%)이라고 해석할 수 있다. [2][3][7]

4. 결론 및 향후 연구

본 논문에서는 공항에서의 입국 심사에, 데이터 마이닝 기법 중 신경망 기법을 적용하여, 효율적이고 기계적인 입국심사가 가능한 방법을 제안하였다. 그러나 국가의 안보가 잠재적인 개인정보침해 등과 대립될 수 있는 것은 분명한 사실이기에, 이 부분에 대해 개인정보보호를 위한 법률이나 방법론이 보완된다면 실전에서 유용하게 사용될 수 있을 것이다.

본 논문에서는 결과에 영향을 미치는 변수에 대해서 가능성 있는 변수들을 직접 선별했지만, 앞으로는 상관분석, 의사결정나무분석 등 다른 데이터 마이닝 기법을 적용하여 사전에 변수들을 선별하고, 그 변수들을 이용하여 향상된 학습을 진행할 것이다. 또한 신경망 기법의 적용 가능한 새로운 분야에 대해 탐구하고, 그 분야에 적합한 신경망의 구조, 변수 값의 변환 및 결과의 재해석 등을 연구할 것이다.

감사의 글

이 논문은 2009 년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2009-0075771).

참고문헌

- [1]http://www.immigration.go.kr/HP/TIMM/imm_06/imm_2010_03.jsp
- [2] 최국렬, 이상렬, 박일수, 김옥남, 강성홍, 조대현, 석경하, 김유미, 김병수 공저, "데이터마이닝 이론과 실습", 청구문화사, 2001
- [3] 조용준, 허준, 최인규 공저, "Neural Connection을 이용한 데이터마이닝 신경망 분석", SPSS 아카데미, 2000
- [4] Randall S.sexton and Nahell A.Sikander "Data Mining Using a Genetic Algorithm-Trained Neural Network", Computer Information Systems, Southwest Missouri State University, USA
- [5] <http://www.statewatch.org/terrorlists/thelists.html>
- [6] http://www.tiic.go.kr/service/info/situatin.do?method=list&manage_cd=001007000
- [7] 김대수, "신경망 이론과 응용", 하이테크정보, 1989