신경망을 이용한 DES 암호분석에 관한 연구

권수진[2], 임형신[2], 강주성[1,2], 염용진[1,2]*
국민대학교 정보보안암호수학과[1] / 금융정보보안학과[2]
{tnwls1595, kuunh2, jskang, *salt}@kookmin.ac.kr

A study on the cryptanalysis of DES using Neural Network

Sujin Kwon[2], Hyoungshin Yim[2], Ju-Sung Kang[1,2], Yongjin Yeom[1,2]*
Dept. of Information Security, Cryptology, and Mathematics[1] /
Financial information security[2], Kookmin Univ.

요 약

최근 암호분석 분야에 신경망을 이용한 연구가 활발하게 진행되고 있다. 대표적으로, WorldCIS 2012에서 Mohammed가 암호키 없이 암호문만으로 DES의 평문을 복호화 가능성을 제시한 연구가 있다. 이 연구에서 사용한 신경망의 근사 가능성은 다층 피드포워드 신경망이 Universal approximator가 된다는 증명에 기반하였다. 하지만, 이산함수인 DES의 근사 가능성은 이 증명에 의해 보장된다는 이론적인 근거가 없다. 따라서, 본 논문에서는 Mohammed가 제안한 신경망을 이용한 DES 암호분석을 검증하기 위해 실험을 재연한다.

Ⅰ. 서 론

인간의 뉴런 구조를 본떠 만든 기계학습 모델인 인공신경망(artificial neural network, ANN)은 관찰된 데이터로부터 학습하여 원하는 근사 함수를 만들 수 있다는 장점으로 다양한 분야에서 응용되고 있다. 암호분석(cryptanalysis) 분야에도 신경망을 적용한 블록암호 공격 기법에 관한 연구가 활발히 진행되고 있다. 대표적으로, 블록암호 DES(data encryption standard)를 암호키 없이 암호문만으로 평문을 복구하는 신경망을 이용한 공격 기법에 관한 연구와[1, 2] 유사한 기법으로 신경망을 이용하여 AES(advanced encryption standard) 평문 복구에 성공하였다는 연구가 있다[3].

본 논문에서는 WorldCIS 2012에서 Mohammed가 제안한 신경망에 기반을 둔 블록암호 DES의 공격 기법 연구[1]에 대한 실험을 재연한다. 재연한 결과에 기반하여 암호키 없이 성공적으로 DES를 복호화한다는 기존연구[1]의 결과를 분석한다.

Ⅱ. 신경망을 이용한 DES 암호분석 선행연구

이 논문에서는 (평문, 암호문) 쌍을 안다는 가정하에, 신경망을 이용해 암호키 없이 ECB(electronic codebook) 모드인 DES 복호화 알고리즘과 기능적으로 같은 알고리즘을 찾는 Global Deduction 공격을 목표로 한다.

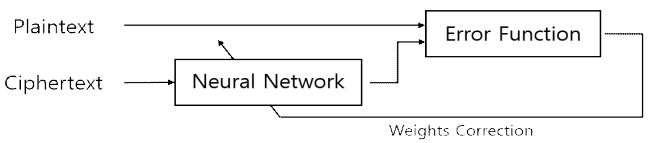

그림 1. 신경망을 이용한 DES 암호분석[1]

선행연구에서 제안한 신경망 기반 DES 암호분석의 신경망 훈련과정(training process)은 그림 1과 같이 진행된다. 암호문은 신경망의 입력이고, 이에 대응하는 평문은 신경망의 입력에 대한 레이블(label)값이다. 훈련 데이터를 사용하여 신경망의 학습을 진행하며, 오차함수(error function)를 통해 신경망의 출력값과 암호문에 대응되는 레이블의 오차를 계산하여 신경망의 가중치와 편향을 조절한다. 훈련과정에서 사용되는 오차함수는 평균제곱오차(mean squared error)를 사용하며, 식은 다음과 같다.

$$\text{평균제곱오차} = \frac{1}{n}\sum_{i=1}^{n}\left|p_i - \hat{p_i}\right|^2.$$

이때, p_i는 훈련 데이터의 레이블값, $\hat{p_i}$는 신경망이 예측한 평문값, n은 훈련 데이터의 개수이다.

신경망의 학습이 종료된 후, 검증과정(validation process)을 진행한다. 훈련된 신경망을 평가하기 위해 검증과정은 내부오차(inside-error)와 외부오차(outside-error)를 측정하여 결과를 제시한다. 내부오차는 훈련에 사용된 암호문으로 예측한 결괏값과 실제 평문값의 오차를 계산한 결과이고, 외부오차는 훈련에 사용되지 않은 암호문으로 예측한 결괏값과 이에 대응되는 평문값의 오차를 계산한 결과이다. 내부오차와 외부오차의 계산법은 다음과 같다.

$$\text{내부오차 및 외부오차} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{t} p'(i,j) \oplus p(i,j)}{s \times t}.$$

이때, s는 사용된 블록의 개수이고, t는 블록의 길이다. $p'(i,j)$는 예측된 평문 i번째 블록의 j번째 비트이며, $p(i,j)$는 실제 평문 i번째 블록의 j번째 비트이다. 선행연구의 저자가 제시한 평균제곱오차, 내부오차, 외부오차의 결과는 표 1에 나타나 있다.

표 1. 신경망을 이용한 DES 암호분석 기존 연구 결과

	평균제곱오차	내부오차	외부오차
오차	0.013317	0.027973	0.085986

선행연구에서는 평균 2^{11}개 미만의 (평문, 암호문) 데이터 쌍을 이용해 DES의 암호문을 성공적으로 복호화할 수 있다고 주장한다.

III. 실험 재연 및 분석

2장의 선행연구는 신경망이 Universal approximator가 된다는 증명에 기반하여 DES 암호분석을 진행한 것으로 보인다. 하지만, 이 증명에서는 DES와 같은 이산함수에 대한 근사 가능성을 다루지 않기 때문에 선행연구가 진행한 실험의 검증이 필요하다. 본 장에서는 선행연구가 진행한 실험을 재연하고, 표 1의 결과와 재연한 실험의 결과를 분석한다.

3.1. 하이퍼 파라미터 및 실험과정

선행연구의 신경망 기반 DES 암호분석은 훈련과정을 모두 마친 후 검증과정을 진행한다. 여기에서의 검증과정은 신경망 모델을 평가하는 테스트 과정과 같다고 판단하였다. 본 논문에서는 훈련과정에서 측정한 평균제곱오차를 훈련오차로 보았고, 훈련에 사용하지 않은 새로운 데이터로 신경망 모델을 평가한 외부오차를 테스트오차로 판단하였다.

선행연구의 실험과정에는 훈련 데이터의 개수와 테스트 데이터의 개수에 대하여 명확하게 제시되어 있지 않다. 따라서 평균 2^{11}개의 (평문, 암호문) 쌍으로 성공적인 실험 결과를 나타내었다는 선행연구의 주장에 기반하여 본 논문에서는 훈련 데이터의 개수를 2^{11}개로 설정하였다. 또한, 명확하게 제시한 파라미터에 대해서는 동일하게 설정하였다.

표 2. 실험 데이터 및 하이퍼 파라미터 설정

훈련 데이터 개수	2,048
테스트 데이터 개수	512
에폭 수	10,000
노드 개수	128(입력)-128-256-256-128(출력)

본 논문에서 재연한 실험의 데이터 개수 및 하이퍼 파라미터(hyper parameter)는 표 2에 나타나 있다. 완전 연결 신경망(fully connected neural network)을 사용하며 활성화 함수는 비선형 함수인 시그모이드(sigmoid)를 사용한다.

본 논문에서 재연한 실험은 다음과 같이 진행된다.

1. 데이터 수집 : 랜덤한 64-비트 평문 두 블록을 추출하여 DES-ECB 모드로 암호화를 하여 64-비트 암호문 두 블록을 생성해 (평문, 암호문) 쌍을 수집한다. 이후 훈련 데이터와 테스트 데이터로 분류한다.
2. 신경망 훈련과정 : 훈련 데이터를 이용해 신경망을 학습시킨다. 오차함수를 통해 신경망이 출력한 예측된 평문과 실제 평문의 오차를 계산하고 가중치와 편향값을 조절한다.
3. 신경망 테스트과정 : 신경망 훈련이 끝난 후, 훈련 데이터와 테스트 데이터로 신경망 모델을 평가한다. 신경망 출력의 각 노드값은 0과 1 사이이므로, 반올림하여 0 또는 1로 만들어 준 뒤 기존 평문과의 오차를 계산한다.

3.2. 실험 결과 및 분석

표 3. 재연한 실험에서 측정된 오차 결과

	훈련오차	내부오차	테스트오차
오차	0.012249	0.160527	0.475657

본 논문에서 재연한 실험의 측정된 오차 결과는 표 3에 나타나 있다. 훈련오차는 선행연구의 실험 결과와 비슷하였다. 한편, 내부오차와 테스트오차는 표 1보다 높은 오차를 가진다.

통상적으로 신경망은 오차가 감소하는 방향으로 학습이 진행되기 때문에, 훈련오차는 낮은값으로 나오는 것은 타당하다. 평균제곱오차는 연속함수에 대한 오차함수이기 때문에, 이산함수인 DES를 학습하는 과정에서 평균제곱오차를 사용하는 것은 잘못된 설정으로 보인다. 내부오차는 0.16으로, 128개의 비트 중 평균 20개의 비트가 잘못 예측되었음을 의미한다. 훈련에 사용한 데이터로 신경망을 평가한 것이므로 테스트오차보다 작은 값으로 계산되는 것은 합당하다. 테스트오차는 약 0.48로, 이는 128개의 비트 중 평균 61개의 비트가 잘못 예측되었음을 의미한다.

선행연구와 본 실험의 훈련된 신경망은 잘못 예측된 비트의 위치에 대한 정보를 주지 않는다. 만약, 신경망의 출력에서 잘못 예측된 비트에 대한 위치정보를 알 수 있다면, 이에 해당하는 비트값을 반전시켜 평문을 복구할 수 있다. 하지만, 잘못 예측된 비트의 위치가 유동적이라면 그림 2와 같이 위치에 대한 전수조사가 필요하다.

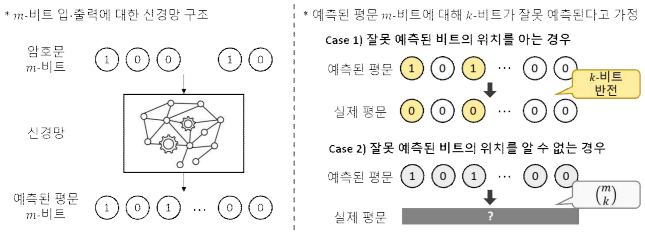

그림 2. 잘못 예측된 비트의 위치정보 제공 여부에 따른 전수조사

표 3을 그림 2와 같이 분석하면, 테스트오차는 $\binom{128}{61}$ 만큼의 틀린 위치에 대한 전수조사가 필요하여 약 2^{123} 정도의 계산량을 가지고, 내부오차의 전수조사량은 $\binom{128}{20} \approx 2^{76}$이다. 그러므로 전체 평문에서 틀린 비트의 비율을 오차로 바라보고 평문을 성공적으로 복구했다는 선행연구의 결과는 타당하지 않다고 분석한다.

IV. 결론

본 논문에서는 신경망을 이용하여 DES의 평문을 복구하는 Mohammed의 연구를 실험적으로 분석하였다. 기존 연구에서 제시된 실험 설정과 유사한 환경에서 재연을 진행하였으며, 재연을 통해 신경망을 이용하여 DES 복호 알고리즘에 근사하였다는 저자의 주장은 신뢰성이 낮음을 확인하였다. 이 논문과 유사한 기법으로 실험을 진행한 다른 블록암호 분석 연구[2-3]에도 이와 같은 문제가 있을 것으로 보인다. 따라서 본 연구진은 추후 연구로 이 논문에서 분석한 공격과 유사한 기법을 사용한 연구를 검토하고, 신경망을 이용한 이산함수의 근사 가능성에 관해 연구할 예정이다.

참 고 문 헌

[1] Alani M. M. " Neuro-cryptanalysis of DES," World Congress on Internet Security (WorldCIS-2012). pp. 23-27, June. 2012.

[2] Fan S., and Zhao Y. "Analysis of DES Plaintext Recovery Based on BP Neural Network." Security and Communication Networks 2019.

[3] Hu X., and Zhao Y. "Research on plaintext restoration of AES based on neural network." Security and Communication Networks 2018.

[4] Hornik K., Stinchcombe M., and White H. "Multilayer feedforward networks are universal approximators," Neural networks 2.5, pp. 359-366.