Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.6, Dec. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.6.861

다양한 차수의 합성 미니맥스 근사 다항식이 완전 동형 암호 상에서의 컨볼루션 신경망 네트워크에 미치는 영향*

이 정 현,[1†] 노 종 선[2‡]
[1,2]서울대학교 (대학원생, 교수)

The Impact of Various Degrees of Composite Minimax Approximate Polynomials on Convolutional Neural Networks over Fully Homomorphic Encryption*

Junghyun Lee,[1†] Jong-Seon No[2‡]
[1,2]Seoul National University (Graduate student, Professor)

요 약

보안을 유지하는 가운데 딥 러닝을 이용하여 데이터 분석 결과를 제공하는 서비스의 핵심적인 기술 중의 하나로 완전 동형 암호가 있다. 완전 동형 암호화된 데이터 간의 연산의 제약으로 인해 딥 러닝에 사용되는 비산술 함수를 다항식으로 근사해야 한다. 현재까지는 합성 미니맥스 다항식을 사용하여 비산술 함수를 근사한 다항식을 컨볼루션 뉴럴 네트워크에 적용했을 때 계층별로 같은 차수의 다항식만 적용하였는데, 이는 완전 동형 암호를 위한 효과적인 네트워크의 설계에 어려움을 준다. 본 연구는 합성 미니맥스 다항식으로 설계한 근사 다항식의 차수를 계층별로 서로 다르게 설정하여도 컨볼루션 뉴럴 네트워크에서 데이터의 분석에 문제가 없음을 이론적으로 증명하였다.

ABSTRACT

One of the key technologies in providing data analysis in the deep learning while maintaining security is fully homomorphic encryption. Due to constraints in operations on fully homomorphically encrypted data, non-arithmetic functions used in deep learning must be approximated by polynomials. Until now, the degrees of approximation polynomials with composite minimax polynomials have been uniformly set across layers, which poses challenges for effective network designs on fully homomorphic encryption. This study theoretically proves that setting different degrees of approximation polynomials constructed by composite minimax polynomial in each layer does not pose any issues in the inference on convolutional neural networks.

Keywords: Fully Homomorphic Encryption, Privacy-Preserving Machine Learning, Composite Minimax Polynomial, Convolutional Neural Network

I. 서 론

클라우드 컴퓨팅 기술의 발전에 따라, 데이터를 분석하는 데 있어 클라이언트와 서버 간의 데이터 통신을 고려하는 방법이 활발히 제안되고 있다. 하지만 보안에 민감한 데이터의 경우 데이터를 전송하는

Received(10. 11. 2023), Modified(11. 10. 2023), Accepted(11. 10. 2023)
* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00400, 저사양 디바이스 대상 고효율 PQC 안전성 및 성능 검증 기술 개발)
† 주저자, jhlee@ccl.snu.ac.kr
‡ 교신저자, jsno@snu.ac.kr(Corresponding author)

데에 한계가 있기 때문에, 이에 대한 접근으로서 완전 동형 암호(fully homomorphic encryption)를 활용한 데이터 분석 방법이 있다[1,2,3]. 완전 동형 암호 기법으로 데이터를 암호화할 경우, 데이터의 정보가 누출될 위험을 방지하면서도 복호화 또는 제 3자의 도움 없이 암호화된 데이터 간의 연산이 가능하기 때문이다. 현재는 완전 동형 암호화된 다양한 데이터를 깊은 신경망 네트워크를 통하여 분석하는 연구가 활발히 진행되고 있다. 이를 보안유지 머신러닝(privacy preserving machine learning: PPML)이라고 한다.

완전 동형 암호는 강력한 보안 유지가 보장되는 대신 수행되는 연산에 많은 제약이 있다. 실수 암호문 간에는 덧셈과 곱셈만 지원될 뿐 아니라, 연속적 곱셈의 가능한 횟수에도 제한이 있다. 따라서 완전 동형 암호화된 데이터를 딥 러닝 모델에서 분석하기가 까다롭다. 딥 러닝 모델 성능 향상에 주요한 기여를 하는 활성화 함수는 대부분 덧셈과 곱셈만으로 표현하기 어려운 비산술 함수이기 때문에 완전 동형 연산만으로 정확하게 구현할 수 없다. 또, 모델의 깊이가 깊은 경우 곱셈하는 데에 문제가 발생한다.

비산술 함수를 다항식으로 대체하기 위해서 크게 두 가지 방향의 연구가 진행되어 왔는데, 비산술 함수를 작은 차수의 다항식으로 근사하여 연산의 시간을 획기적으로 줄이는 방법[1], 그리고 비산술 함수의 근사 오차를 최대한 줄여 정확도를 높이는 방법이 있다[2,3,4]. 첫 번째의 방법은 깊은 네트워크와 큰 데이터셋에 적용하기 어렵기 때문에[5], 깊은 네트워크에 동형 암호화된 데이터를 적용하기 위해서 정확도를 최대한 높이면서 비산술 함수를 근사하는 두 번째 방법이 중요하게 여겨지고 있다.

비산술 함수를 근사하는 방법은 근사 오차의 측정 방식에 따라 달라진다. 합성 미니맥스 다항식(composite minimax polynomial)[6]을 사용하여 근사 영역에서 근사 오차를 균등하게 제한하는 방법[2,3,4], 근사 영역에서 입력값의 분포를 고려하여 상대적으로 입력값이 밀집된 곳에서 근사 오차를 더욱 제한하는 방법[7]이 있다.

합성 미니맥스 다항식은 ReLU 함수와 맥스-풀링(max-pooling) 함수를 임의의 오차 이내로 정확하게 근사할 수 있도록 하고, ResNet과 VGGNet 등의 컨볼루션 뉴럴 네트워크(convolutional neural network: CNN)에 잘 적용되는 것이 밝혀져 있다[5]. 하지만, 이 다항식은 지금까지의 연구 결과로는 CIFAR-10 데이터셋의 이미지 한 장의 분석(inference)을 위해 2,000초 이상의 시간이 필요할 정도로 굉장히 많은 연산 시간을 요구한다[2,3]. 소요되는 시간의 대부분은 근사 다항식의 높은 차수로 인해 발생한다. 다항식 자체의 연산 시간뿐 아니라, 높은 차수의 다항식 연산을 통해 여러 번 곱셈을 수행하고 나면 더 이상 동형 연산을 수행할 수 없게 되는 상태를 극복하기 위한 부트스트래핑(bootstrapping) 연산까지 고려했을 때 전체 분석 시간의 70% 이상이 높은 다항식 차수로 인하여 소요된다[3]. 합성 미니맥스 다항식을 사용한 근사 방법에서는 각 계층에서의 근사 다항식의 차수의 뚜렷한 결정 기준이 없었고, 각 계층의 다항식 차수를 모두 같게 설정하는 상황에서 최적의 근사 다항식을 찾아 이를 암호문 분석에 사용하였다[5].

그런데, 입력값의 분포를 고려하여 다항식을 근사한 연구 [7]에서는 각 계층의 다항식 차수를 다르게 설정하는 상황까지 폭넓게 고려하였다. Lee 등[7]은 정밀한 근사가 비교적 덜 요구되는 계층에서 낮은 차수의 다항식을 사용하여 각 계층의 부트스트래핑 및 다항식 연산 시간이 최대 50%가 줄일 수 있었다.

본 연구는, 각 계층의 차수를 다양화하는 방법이 합성 미니맥스 다항식을 사용하는 경우에 대해서도 동일하게 적용될 가능성에 대해 논의한 것이다. 입력값의 분포를 고려한 근사 방법은 요구되는 차수가 합성 미니맥스 다항식보다 적다는 장점이 있지만, 다항식의 계수가 매우 큰 경우가 많아 암호문 상에서의 연산이 불안정하고 근사 다항식을 변형시켜 주어야 한다는 불편함이 있다[7]. 따라서 합성 미니맥스 다항식을 사용하는 컨볼루션 뉴럴 네트워크에서도 계층별 차수의 다양화를 시도해도 유의미한 분석이 수행될 수 있음을 밝히는 것이 필요하다. 본 연구에서는 일반적인 컨볼루션 뉴럴 네트워크에서 대부분 사용되는 컨볼루션 연산, ReLU 연산, 맥스-풀링 연산으로 이루어진 임의의 네트워크에서 비산술 함수를 다양한 차수로 근사하여도, 기존 모델의 출력값과 근사 모델의 출력값의 차이가 충분히 줄어든다는 것을 이론적으로 증명하였다.

II. 배경 이론

2.1 완전 동형 암호

완전 동형 암호는 정수, 실수, 또는 복소수 평문

을 암호화한 암호문들 간의 동형 연산을 지원하여 암호문 간의 의미 있는 연산을 수행할 수 있도록 한다. 본 연구에서는 평문 공간(message space)을 실수 벡터 공간 R^N으로 정의하여 실제 상황에서 얻어지는 데이터들을 모두 암호화할 수 있도록 한다.

데이터를 암호화하는 사용자는 사용자만의 비밀키 k를 사용하여 평문 $m \in R^N$을 암호화할 수 있는 암호화 함수 $Enc_k(m)$, 그리고 비밀키 k로 암호화된 암호문 c를 복호화할 수 있는 복호화 함수 $Dec_k(c)$를 사용할 수 있다. 이 함수들은 비밀키 k가 없이는 수행될 수 없는 기능들이다.

반면, 암호문 간의 동형 연산은 비밀키를 가지고 있지 않은 제 3자도 수행할 수 있다. 완전 동형 암호에서는 다음의 동형 연산들을 지원한다.

- ▷ 동형 덧셈 (homomorphic addition, $\oplus$):
  - $Enc_k(m_1) \oplus Enc_k(m_2) = Enc_k(m_1 + m_2)$
- ▷ 동형 스칼라 곱셈 (homomorphic scalar multiplication, $\odot$):
  - $Enc_k(m) \odot r = Enc_k(m \cdot r)$
- ▷ 동형 넌스칼라 곱셈 (homomorphic non-scalar multiplication, $\otimes$):
  - $Enc_k(m_1) \otimes Enc_k(m_2) = Enc_k(m_1 \times m_2)$

위의 두 연산을 통해 비밀키가 없는 사용자도 덧셈과 곱셈으로 이루어진 산술 연산을 진행할 수 있으며, 최종 연산 결과의 평문은 비밀키를 가진 사용자가 최종 암호문을 복호화함으로서 얻어낼 수 있다.

동형 연산의 종류에는 위와 같이 제약이 있으므로, 비산술 함수의 연산 시에는 덧셈과 곱셈만으로 이루어진 다항식 연산으로 대체되어야 한다.

또한, 암호문 간의 연산에서는 특별히 곱셈의 횟수에도 제약이 있다. 곱셈을 일정 횟수 이상 수행하게 되면 암호문의 모듈러스(modulus)가 줄어들며 암호문 간의 곱셈이 더 이상 불가능해지는 시점이 존재한다. 곱셈을 더 이상 수행할 수 없는 암호문에 부트스트래핑을 적용하면 모듈러스가 일정 값만큼 증가하며 다시 곱셈을 수행할 수 있지만, 부트스트래핑의 경우는 다른 연산 대비 많은 시간을 소모한다[3]. 따라서 암호문 간의 동형 연산이 많이 필요한 경우에는 적절한 최적화 작업이 필요하다.

2.2 합성 미니맥스 다항식

컨볼루션 뉴럴 네트워크에서 가장 많이 활용되는 비산술 함수로서 ReLU 함수와 맥스-풀링 함수가 있다. Lee 등[5]은 ReLU 함수와 최댓값 함수를

$$ReLU(x) = \frac{x}{2}(1 + sgn(x))$$
$$\max\{a,b\} = \frac{1}{2}(a + b + (a-b)sgn(a-b))$$

로 표현하여 $sgn(x)$의 근사 다항식을 통해 ReLU 함수 및 최댓값 함수의 근사다항식을 얻어내었다. 여기에서 $sgn(x)$는 부호 함수로서, $x > 0$일 경우 1, $x < 0$일 경우 -1, $x = 0$일 경우 0을 나타내는 함수이다. 부호 함수 $sgn(x)$는 동형 연산에 최적화된 방법으로 다항식으로 근사하는 방법이 연구된 바 있다[6]. 이 방법은 미니맥스 방식으로 근사된 다항식들을 합성함을 통해 동형 곱셈으로 인한 모듈러스의 감소 곧 암호문의 뎁스(depth)의 소모를 최소화하면서 다항식 연산을 수행하는 방법을 제공한다. 이 근사 다항식의 오차는 정확도 파라미터(precision parameter)에 따라 달라지며, 또한 이에 맞게 근사 다항식 연산이 소모하는 뎁스의 수가 달라진다. 정확도 파라미터는 임의의 자연수의 값으로 설정할 수 있으며, 일반적인 컨볼루션 뉴럴 네트워크에서는 α의 값을 13 또는 14의 값으로 설정하면 충분히 높은 성능을 달성할 수 있다[6].

주어진 정확도 파라미터 α에 대해 부호 함수를 뎁스 소모 면에서 최적으로 근사한 다항식을 $p_\alpha(x)$로 쓰도록 한다. 본 연구에서는 부호 함수의 근사 다항식 $p_\alpha(x)$를 $sgn(x)$ 대신 대입하는 방법을 통해 연구 [5]에서 사용된 다음의 조건

$$|ReLU(x) - r_\alpha(x)| \le 2^{-\alpha} \text{ for } x \in [-1,1] \quad (1)$$

$$|M_{\alpha,n}(x_1,\cdots,x_n) - \max\{x_1,\cdots,x_n\}| \le 2^{-\alpha} \lceil \log_2 n \rceil$$
$$\text{for } x_i \in [\epsilon, 1-\epsilon],\ \epsilon = (\lceil log_2 n \rceil - 1)2^{-\alpha} \quad (2)$$

를 만족하는 ReLU 함수의 근사 다항식 $r_\alpha(x)$과 맥스-풀링 함수의 근사 다항식 $M_{\alpha,n}(x_1,\cdots,x_n)$를 사용한다. 등식 (1)과 (2)를 통해 정확도 파라미터 α가 커질수록 비산술 함수의 근사 오차가 줄어듦을 확인

할 수 있다.

연구 [5]에서는 등식 (1)과 (2)에 나타난 제한된 근사영역을 더 넓힐 수 있도록 다항식의 근사 영역 파라미터 B를 정확도 파라미터와 독립적으로 설정하였다. 각각의 비산술 함수에 대해 다항식의 근사 영역을 $[-B,B]$로 확장하고 싶다고 할 때, ReLU 함수의 경우 $\tilde{r}_{\alpha,B}(x) := Br_{\alpha}(x/B)$를 근사다항식으로 사용하고, 맥스-풀링 함수의 경우

$$B' = B/(0.5-(\lceil log_2 n \rceil -1)2^{-\alpha})$$

에 대하여

$$\widetilde{M}_{\alpha,n,B}(x_1,\cdots,x_n) := B' \cdot \left(M_{\alpha,n}(\frac{x_1}{B'}+0.5,\cdots,\frac{x_n}{B'}+0.5)-0.5\right)$$

를 근사다항식으로 사용한다. 위의 정의는 B'의 값이 양수로서 잘 정의될 때에만 유효하다. 각각의 다항식은 다음의 오차 관계를 만족한다[5].

$$|\tilde{r}_{\alpha,B}(x)-ReLU(x)| \le B \cdot 2^{-\alpha},$$
$$|\widetilde{M}_{\alpha,n,B}(x_1,\cdots,x_n) - \max\{x_1,\cdots,x_n\}| \le B'2^{-\alpha}\lceil \log_2 n \rceil .$$

Table 1에는 정확도 파라미터 α에 따라 부호 함수 $sgn(x)$를 근사한 다항식 $p_{\alpha}(x)$의 연산이 소모하는 뎁스를 구체적으로 나타내었다.

Table 1에서 볼 수 있듯이, 정확도 파라미터가 높을수록 다항식 $p_{\alpha}(x)$에서 소모되는 뎁스의 수가 늘어난다. 뎁스의 수를 줄일수록 다항식의 연산 시간뿐 아니라 부트스트래핑이 요구되는 횟수도 줄기 때문에, 이에 따라 딥 러닝 모델 전체의 소요 시간에 영향을 크게 줄일 수 있다. 그러므로 비산술 함수를 다항식으로 근사할 때에는 시간 소모 측면에서 볼 때 가능한 한 최소의 차수로 근사하는 것이 좋다. 반면, 식 (1)에서 볼 수 있듯이 정확도 파라미터가 클수록, 곧 뎁스 소모가 많을수록 비산술 함수들을 정확하게 연산할 수 있다. 즉 정확도 파라미터를 어떻게 정하는가에 따른 시간 소모와 정확도 간의 상충 관계가 존재한다.

Table 1. The depth consumption of the optimal composite minimax polynomial $p_{\alpha}(x)$ with respect to the precision parameters α.

α	4	5	6	7	8	9
depth	3	4	5	6	7	8
α	10	11	12	13	14	15
depth	10	11	12	13	14	15

2.3 보안 유지 머신 러닝

완전 동형 암호를 이용하면, 컨볼루션 뉴럴 네트워크를 암호화된 데이터에 적용하여 데이터를 분석하는 것이 가능하다.

컨볼루션 뉴럴 네트워크를 하나의 함수 $F(x)$로 생각하자. 함수 F는 입력값 x를 받아 출력값 $F(x)$를 내고, $F(x)$의 정보를 활용하여 입력값 x에 대한 속성을 추론한다. 예를 들어 CIFAR-10 데이터셋의 이미지 분류를 위한 컨볼루션 뉴럴 네트워크의 경우 $F(x)$는 길이가 10인 벡터로서 $F(x)$의 값 중 가장 큰 값의 레이블을 통해 이미지의 종류를 예측한다.

보안 유지 머신 러닝의 목표는 입력값 x를 데이터의 소유자가 암호화하여 모델의 소유자에게 전달했을 때 모델의 소유자는 이 암호문을 토대로 $F(x)$의 값을 평문으로 담고 있는 암호문을 만드는 것이다. 즉, 데이터의 소유자가 암호문 $Enc_k(x)$를 모델의 소유자에게 전달하면, 모델의 소유자는 이를 통해 암호문 $Enc_k(F(x))$를 생성하는 것이다. 이 암호문을 데이터의 소유자에게 전달하면, 데이터의 소유자는 비밀키 k를 통해 암호문 $Enc_k(F(x))$를 복호화함으로서 출력값의 평문 $F(x)$를 얻어낼 수 있다.

하지만, 대부분의 컨볼루션 뉴럴 네트워크 함수 $F(x)$의 경우 덧셈과 곱셈만으로 표현하지 못한다. 우리는 $F(x)$를 구성하는 연산들 중 비산술 함수를 동형 암호 연산으로 구현할 수 있도록 다항식으로 근사한 새로운 함수를 $F^{\alpha}(x)$로 표현한다. 여기에서 α는 정확도 파라미터로서 다항식을 얼마나 정확하게 근사하였는지 나타내는 값이다. 예를 들어 ReLU 함수의 경우 식 (1)을 만족하는 다항식 $r_{\alpha}(x)$로 표현할 수 있다. 앞으로는 모델 $F(x)$를 기존 모델, $F^{\alpha}(x)$를 근사 모델이라고 부르기로 한다.

이제 $F^{\alpha}(x)$에 포함된 평문 상의 덧셈과 곱셈을 암호문 상의 덧셈과 곱셈으로 치환한 암호문 상의 함수를 $\overline{F^{\alpha}}(c)$라 한다. 그러면 동형 연산의 성질에 따라 다음의 등식이 성립한다.

$$\overline{F^{\alpha}}(Enc_k(x)) = Enc_k(F^{\alpha}(x)) \quad (3)$$

따라서, 모델의 소유자가 $\overline{F^{\alpha}}(Enc_k(x))$를 출력하여 데이터의 소유자에게 전달하면, 데이터의 소유자는 이를 비밀키 k로 복호화할 수 있고 식 (3)에 의해 그 결과인 $F^{\alpha}(x)$를 얻는다. 만일 기존 딥 러닝 모델의 출력값 $F(x)$와 비산술 함수를 다항식으로 근사한 새로운 함수 $F^{\alpha}(x)$의 값에 큰 차이가 없다면, 우리는 새로운 딥 러닝 모델 $\overline{F^{\alpha}}$를 구현함으로서 데이터 x의 정보를 모델의 소유자에게 누출하지 않으면서 데이터의 소유자가 원하는 $F(x)$의 근삿값을 얻을 수 있게 된다. Lee 등[5]은 합성 미니맥스 다항식을 통하여 ReLU 함수와 맥스-풀링 함수를 근사하였을 때 다음의 등식이 성립하는 상수 C가 존재함을 증명하였다.

$$\| F^{\alpha}(x) - F(x) \|_{\infty} \leq C \cdot 2^{-\alpha} \quad (4)$$

식 (4)를 통해, 정확도 파라미터를 충분히 증가시켰을 경우 근사 모델의 출력값 $F^{\alpha}(x)$를 기존 모델의 출력값 $F(x)$에 충분히 가까워짐을 확인할 수 있다.

기존의 출력값 $F(x)$와 근사 다항식을 적용한 출력값 $F^{\alpha}(x)$의 차이는 컨볼루션 뉴럴 네트워크의 각 계층에 위치한 다항식이 어떻게 근사되어 있는지에 따라 결정된다. 컨볼루션 뉴럴 네트워크 내부의 각각의 연산 단위들을 블록(block)이라고 하자. 그러면 음이 아닌 오차 e에 대해 각각의 블록이 근사된 정도에 따라 발생하는 오차의 전파 함수(error propagation function)

$$E_A^{\alpha}(e) := \sup_{\|x+\vec{e}\|_{\infty} \leq B,\ \|\vec{e}\|_{\infty} \leq e} \|A^{\alpha}(x+\vec{e}) - A(x)\|_{\infty}$$

를 정의할 수 있다[5]. 여기에서 A는 블록, e는 입력값 x에 대해 발생하는 오차들의 벡터 $\vec{e}$의 크기, B는 다항식의 근사 영역 파라미터이다. 각 블록의 오차의 전파 함수의 형태를 파악하면, 비산술 함수를 다항식으로 근사함으로서 발생하는 전체 컨볼루션 뉴럴 네트워크의 분석 결과의 오차에 관련된 정보를 얻을 수 있다.

컨볼루션 뉴럴 네트워크 상에서 자주 사용되는 선형 블록과 ReLU 함수 블록 및 맥스-풀링 함수 블록에 대해서는 오차의 전파 함수가 다음과 같음이 알려져 있다.

정리 1. [5] (a) 블록 A가 선형 블록이고 $A(x) = Wx + b$ 꼴로 나타난다고 할 때, $E_A^{\alpha}(e) \leq \|W\|_{\infty} e$이다. (b) 블록 A가 ReLU 블록, 즉 $A(x) = ReLU(x)$라 하고 다항식 근사 영역이 $[-B,B]$일 때, $E_A^{\alpha}(e) \leq B2^{-\alpha} + e$이다. (c) 블록 A가 $K_0 \times K_0$의 커널(kernel) 크기를 갖는 맥스-풀링 블록이라 하고 다항식 근사 영역이 $[-B,B]$이라 하자. $K_0 \leq 10$이고 $\alpha \geq 4$이면

$$E_A^{\alpha}(e) \leq 10B \lceil log_2 K_0^2 \rceil + e$$

이다.

본 연구에서는 기존의 연구에서 밝혀졌던 각 블록에 대한 오차의 전파 함수를 활용하여, 모델 $F(x)$의 각 계층의 정확도 파라미터가 서로 다르더라도 모델의 동작에는 문제가 없음을 이론적으로 보였다.

III. 합성 미니맥스 다항식의 계층별 근사 변화의 적용 가능성

기존의 합성 미니맥스 근사 다항식의 연구에서는 에 비산술 함수들을 같은 정확도 파라미터를 사용하여 근사하였다. 하지만 본 연구에서는 합성 미니맥스 근사 다항식을 사용하는 컨볼루션 뉴럴 네트워크에서 각각의 블록의 정확도 파라미터에 자유도를 주는 상황을 새롭게 고려한다. 즉, i번째 블록의 비산술 함수의 정확도 파라미터를 α_i로 하여 근사하는 상황을 제안한다. 각각의 블록별로 정확도 파라미터를 $\vec{\alpha} = (\alpha_1, \cdots, \alpha_n)$로 서로 달리 했을 때의 근사 모델을 $F^{\vec{\alpha}}(x)$라고 정의한다. 이 때, 기존 모델의 출력값 $F(x)$와 근사 모델의 출력값 $F^{\vec{\alpha}}(x)$의 차이가 아래와 같이 나타남을 증명할 수 있다.

정리 2. 컨볼루션 뉴럴 네트워크 F가 선형 블록과 ReLU 블록 그리고 맥스-풀링 블록의 합성 $A_n \circ \cdots \circ A_1$로 분해된다고 하자. 이 때, F의 비산술 함수들을 $\alpha_i \geq 4$의 정확도 파라미터로 근사한다고 할 때,

$$\|\vec{F^{\alpha}}(x)-F(x)\|_{\infty} \le \sum_{i=1}^{n} C_i 2^{-\alpha_i} \quad (5)$$

를 만족하는 $\vec{\alpha}=(\alpha_1,\cdots,\alpha_n)$과 무관한 상수 $C_1,\cdots,C_n$이 존재한다.

증명. 증명에 앞서, 블록 A_i가 비산술 함수일 경우 다항식 근사 영역을 $[-B_i,B_i]$라 하고, 블록 A_i가 맥스-풀링 블록일 경우 커널의 크기가 $K_i \times K_i$라 하자. 이 때, 근사 영역 파라미터 B_i와 커널의 크기 K_i들은 정확도 파라미터 $\vec{\alpha}$와 무관한 값이다. 또한, 대부분의 딥 러닝 모델에서 사용되는 맥스-풀링 함수의 커널의 크기는 10을 넘지 않으므로 맥스-풀링 함수의 모든 커널의 크기가 $K_i \le 10$을 만족한다고 가정하자.

블록들의 수 n에 대한 수학적 귀납법으로 등식을 증명한다. 먼저, 블록의 수가 $n=1$일 때를 증명한다. 만약 주어진 블록이 선형 블록일 경우, 근사해야 할 비산술 함수가 없으므로 등식 (5)의 좌변은 0이 된다. 만약 주어진 블록이 ReLU 블록일 경우, 등식 (1)에 의해 좌변은 $B_1 2^{-\alpha_1}$ 이하가 되어 등식 (5)를 만족하는 상수 $C_1 = B_1$가 존재한다. 만약 주어진 블록이 맥스-풀링 블록일 경우, 등식 (2)에 의해 좌변은 $B_1 2^{-\alpha_1}\lceil \log_2 K_1^2 \rceil$ 이하가 되어 등식 (5)를 만족하는 상수 $C_1 = B_1 \lceil \log_2 K_1^2 \rceil$이 존재한다. 따라서, $n=1$일 때의 명제가 성립한다.

이제 블록의 개수가 $n=k$일 때 명제가 성립한다고 가정하자. 그리고 블록의 개수가 $n=k+1$일 때 명제가 성립함을 확인하자. 먼저, 블록 A_1부터 A_k까지를 합성한 모델 F_k에 대해 귀납 가정을 적용하면,

$$\|\vec{F_k^{\alpha'}}(x)-F_k(x)\|_{\infty} \le \sum_{i=1}^{k} C_i 2^{-\alpha_i}$$

가 성립하는 상수 $C_1,\cdots,C_k$가 존재한다. 여기에서 $\vec{\alpha'}=(\alpha_1,\cdots,\alpha_k)$이다.

먼저 임의의 오차의 전파 함수는 항상 단조증가함수임을 언급한다. 음이 아닌 실수 $0 \le e_1 < e_2$에 대해 $\|x+\vec{e}\|_{\infty} \le B$와 $\|\vec{e}\| \le e_1$를 동시에 만족하도록 하는 x와 $\vec{e}$는 항상 $\|x+\vec{e}\|_{\infty} \le B$와 $\|\vec{e}\|_{\infty} \le e_2$ 역시 동시에 만족하므로, 오차의 전파 함수는 단조증가함수이다. 따라서, $E_{A_{k+1}}^{\alpha_{k+1}}(\cdot)$의 단조성에 의해

$$E_{A_{k+1}}^{\alpha_{k+1}}(\|\vec{F_k^{\alpha'}}(x)-F_k(x)\|_{\infty}) \le E_{A_{k+1}}^{\alpha_{k+1}}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right)$$

가 성립한다. 정리하면,

$$\begin{aligned}
&\|\vec{F_{k+1}^{\alpha}}(x)-F_{k+1}(x)\|_{\infty} \\
&= \|A_{k+1}^{\alpha_{k+1}}(\vec{F_k^{\alpha'}}(x))-A_{k+1}(F_k(x))\|_{\infty} \\
&\le E_{A_{k+1}}^{\alpha_{k+1}}(\|\vec{F_k^{\alpha'}}(x)-F_k(x)\|_{\infty}) \\
&\le E_{A_{k+1}}^{\alpha_{k+1}}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right)
\end{aligned}$$

가 성립한다.

만약 A_{k+1}이 컨볼루션 블록 $Wx+b$라면, 정리 1(a)로부터

$$E_{A_{k+1}}^{\alpha_{k+1}}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) \le \|W\|_{\infty}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) = \sum_{i=1}^{k+1} C_i' 2^{-\alpha_i}$$

가 된다. 여기에서 $i=1,\cdots,k$에 대해 $C_i' = \|W\|_{\infty} C_i$이고 $C_{k+1}' = 0$이다. 만약 A_{k+1}이 ReLU 블록이라면, 정리 1(b)로부터

$$\begin{aligned}
E_{A_{k+1}}^{\alpha_{k+1}}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) &\le B_{k+1} 2^{-\alpha_{k+1}} + \left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) \\
&= \sum_{i=1}^{k+1} C_i' 2^{-\alpha_i}
\end{aligned}$$

가 된다. 여기에서 $i=1,\cdots,k$에 대해 $C_i' = C_i$, $C_{k+1}' = B_{k+1}$가 된다. 만약 A_{k+1}이 맥스-풀링 블록이라면, 정리 1(c)로부터

$$\begin{aligned}
E_{A_{k+1}}^{\alpha_{k+1}}\left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) \le\ & 10 B_{k+1} 2^{-\alpha_{k+1}} \lceil \log_2 K_{k+1}^2 \rceil \\
&+ \left(\sum_{i=1}^{k} C_i 2^{-\alpha_i}\right) = \sum_{i=1}^{k+1} C_i' 2^{-\alpha_i}
\end{aligned}$$

가 된다. 여기에서 $i=1,\cdots,k$에 대해 $C_i' = C_i$, $C_{k+1}' = 10 B_{k+1} \lceil \log_2 K_{k+1}^2 \rceil$가 된다. 각각의 경우

에 대해 $n=k+1$일 때 등식 (5)를 만족하는 $\vec{\alpha}$와 무관한 상수 C_i'가 존재하므로, 본 정리가 $n=k+1$일 때도 성립한다.

따라서, n에 대한 수학적 귀납법에 의해 본 정리가 성립한다. □

정리 2를 통해, 다항식으로 근사할 각각의 비산술 함수의 정확도 파라미터가 달라도 각 정확도 파라미터에 대응되는 오차 $2^{-\alpha_i}$들을 충분히 감소시키면 기존 모델의 출력값 $F(x)$와 근사 모델의 출력값 $F^{\vec{\alpha}}(x)$의 차 역시 충분히 감소함을 확인할 수 있다. 또한, 두 출력값의 차이의 크기는 각 비산술 함수의 근사 오차 $2^{-\alpha_i}$의 선형결합보다 작게 나타남을 확인할 수 있다.

IV. 결 론

본 연구에서는 완전 동형 암호화된 데이터의 이미지 분류를 위하여 컨볼루션 뉴럴 네트워크에 사용되는 비산술 함수 중 합성 미니맥스 다항식으로 근사되는 ReLU 함수와 맥스-풀링 함수들에 대해, 같은 차수의 다항식들로 비산술 함수를 근사하는 기존의 연구와 다르게 서로 다른 차수의 다항식으로 근사하는 가능성을 새롭게 고려하였다. 그리고 서로 다른 차수의 다항식으로 근사하여도 기존 모델의 출력값과 근사 모델의 출력값의 차이가 충분히 줄어들 수 있음을 이론적으로 증명하였다. 본 연구에서 새롭게 증명한 정리를 통해, 동형암호를 사용하는 컨볼루션 뉴럴 네트워크에서 비산술 함수의 근사 다항식의 차수가 계층별로 다르게 적용되어도 암호화된 데이터를 좋은 성능으로 분류할 수 있음을 확인할 수 있었다.

본 연구에서는 ReLU 함수와 맥스-풀링 함수에 한정된 비산술 함수에 대하여 논의가 진행되었다. 하지만 딥 러닝 모델에 사용되는 다른 임의의 비산술 함수와 특정 근사 영역이 주어졌을 때 다항식 근사를 수행할 수 있는 방법이 제안된다면, 여기에 대한 오차의 전파 함수를 새롭게 얻어 정리 2의 부등식과 유사한 결과를 얻을 수 있을 것이다. 또한, 본 연구에서는 컨볼루션 뉴럴 네트워크의 분석에 대한 다항식 근사에 대한 논의만 진행되었지만, 컨볼루션 뉴럴 네트워크의 학습 과정을 분석하여 완전 동형 암호화된 데이터로 비산술 함수가 포함된 네트워크를 학습할 수 있는지에 대한 연구도 추가로 진행될 수 있을 것으로 여겨진다.

References

[1] F. Boemer, Y. Lao, R. Cammarota, and C. Wierzynski, "NGraph-HE: A graph compiler for deep learning on homomorphically encrypted data," in Proceedings of 16th ACM International Conference on Computing Frontiers, pp. 3 - 13, Apr. 2019.

[2] J. Lee, H. Kang, Y. Lee, W. Choi, J. Eom, M. Deryabin, E. Lee, J. Lee, D. Yoo, Y. Kim, and J. No, "Privacy-preserving machine learning with fully homomorphic encryption for deep neural network," IEEE Access, vol. 10, pp. 30039-30054, Mar. 2022.

[3] E. Lee, J. Lee, J. Lee, Y. Kim, Y. Kim, J. No, and W. Choi, "Low-complexity deep convolutional neural networks on fully homomorphic encryption using multiplexed parallel convolutions," Proceedings of International Conference on Machine Learning 2022, pp. 12403-12422, June. 2022.

[4] D. Kim, and C. Guyot. "Optimized privacy-preserving CNN inference with fully homomorphic encryption," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 2175 - 2187, Mar. 2023.

[5] J. Lee, E. Lee, J. Lee, Y. Kim, Y. Kim, and J. No, "Precise approximation of convolutional neural networks for homomorphically encrypted data," IEEE Access, vol. 11, pp. 62062 - 62076, June. 2023.

[6] E. Lee, J. Lee, J. No, and Y. Kim, "Minimax approximation of sign function by composite polynomial for homomorphic comparison," IEEE Transactions on Dependable and

Secure Computing, vol. 19, no. 6, pp. 3711-3727, Aug. 2021.
〔7〕 J. Lee, E. Lee, Y. Kim, Y. Lee, J. Lee, Y. Kim, and J. No, "Optimizing layerwise polynomial approximation for efficient private inference on fully homomorphic encryption: a dynamic programming approach," arXiv preprint arXiv:2310.10349, Oct. 2023.

〈저자소개〉

이 정 현 (Junghyun Lee) 학생회원
2018년 2월: 서울대학교 통계학부 졸업
2021년 2월: 서울대학교 수리과학부 석사
2021년 3월~현재: 서울대학교 전기전자공학부 박사과정
〈관심분야〉 정보보호, 암호학, 인공지능

노 종 선 (Jong-Seon No) 종신회원
1981년 2월: 서울대학교 전자공학과 졸업
1984년 2월: 서울대학교 대학원 전자공학과 석사
1988년 2월: University of Southern California, Department of Electrical Engineering, 박사
1999년 8월~현재: 서울대학교 전기정보공학부 교수
〈관심분야〉 암호학, 오류정정부호, 인공지능 보안