완전 동형 암호화된 데이터를 위한 근사 합성곱 신경망의 근사 범위에 따른 성능에 관한 연구

이정현, 노종선*

서울대학교

jhlee@ccl.snu.ac.kr, *jsno@snu.ac.kr

A Study on the Performance of the Approximate Convolutional Neural Network According to Approximation Bound for Homomorphically Encrypted Data

Lee Junghyun, No Jong-Seon*

Seoul National Univ.

요 약

클라우딩 컴퓨팅의 보안을 유지하면서도 데이터에 대한 분석을 진행하기 위해 완전 동형 암호를 사용한 딥 러닝 모델이 최근에 활발하게 연구되고 있다. 딥 러닝 모델이 완전 동형 암호에 적용되기 위해서는 다항식으로 연산이 불가능한 함수들을 다항식으로 근사하는 것이 필요하다. 본 논문의 목표는 이러한 함수를 다항식으로 근사할 때 성능에 어떤 변화가 일어나는지 알아보는 것을 목표로 하였다. 특히, 본 논문의 시뮬레이션을 통하여 ReLU 함수의 다항식 근사 범위 설정이 기학습된 모델이 성능에 중요한 역할을 하고 있음을 확인할 수 있었다.

Ⅰ. 서 론

딥 러닝은 지속적인 기술 발전에 따라 데이터 분석을 위한 필수적인 방법이 되었고, 클라우딩 컴퓨팅을 활용하여 다양한 사람들에게 딥 러닝 모델을 활용한 데이터 처리 서비스를 진행하기도 한다. 근래에는 금융, CCTV 자료, 환자 의료정보와 같은 보안 유지가 필수적인 데이터의 분석에 딥 러닝을 활용기도 한다. 데이터를 가지고 있는 고객이 딥 러닝을 모델을 갖고 있는 다른 서버에게 데이터 분석을 의뢰한다고 할 때, 고객이 가지고 있는 데이터의 보안이 유지되어야 한다면 쉽게 데이터를 서버에게 전달할 수 없을 것이다. 이를 위해 고객은 데이터를 암호화하여 서버에게 전달해야 하는데, 암호화가 되어 있는 채로 데이터를 전달받은 서버는 복호화의 과정 없이 데이터 간의 원하는 연산을 진행하기가 힘들다. 이러한 문제를 해결할 수 있는 암호화 스킴이 동형 암호이다.

동형 암호(Homomorphic Encryption)는 암호화된 데이터 간의 연산, 덧셈과 곱셈을 지원하는 암호 스킴인데, 이 연산에는 특별한 복호화 과정이 필요없다는 것이 큰 장점이다. 따라서 동형 암호 스킴으로 고객의 데이터를 암호화할 수 있다면 고객의 데이터의 정보가 노출되지 않으면서도 서버가 원하는 연산을 수행하여 보안성이 유지되는 데이터 처리를 할 수 있게 되는 것이다. 동형 암호 스킴은 [1]에서 처음으로 격자 문제에 기반하여 제안되었고 그 이후로 동형 암호의 효율성 개선에 대한 연구가 활발히 진행되어 왔다. 최근 [2]에서 제안한 Homomorphic Encryption for Arithmetic of Approximate Numbers (HEAAN 또는 CKKS)인데, 이는 실수 연산에 적합하다는 특징이 있어 실제 데이터 분석에 활용 가능성이 높을 것으로 주목되고 있다.

동형 암호를 사용하여 데이터를 처리하는 방법에는 크게 두 가지가 있다. 첫째는 완전 동형 암호 (Fully Homomorphic Encryption) 기법으로서, 다른 도움 없이 서버가 제3자의 도움 없이 처음부터 끝까지 모든 연산을 할 수 있도록 하는 것이다. 이를 위해서 서버는 동형 암호에서 제한적으로 제공하는 덧셈과 곱셈만으로 데이터 처리를 진행해야 한다. 둘째는 부분 동형 암호 (Somewhat Homomorphic Encryption) 기법으로서 다자간 연산(Multi-Party Computation) 방법을 사용하여 보다 넓은 범위의 연산을 가능하도록 하는 것이다. 부분 동형 암호의 큰 장점은 서버가 덧셈과 곱셈 이외의 연산을 비교적 자유롭게 할 수 있다는 것이지만, 이를 위해서는 특정 데이터를 고객과 주고받아야 한다. 이 과정에서 서버는 데이터 통신에 대한 비용뿐 아니라 자신이 갖고 있는 데이터 처리 기술(예를 들면 딥 러닝 모델의 구조)의 유출에 대한 위험을 감수해야 한다. 반면, 완전 동형 암호를 사용할 경우 데이터 처리에 대한 보안은 서버가 완벽하게 유지할 수 있다. 대신 아무리 간단한 연산이라도 덧셈과 곱셈이 아니라면 이를 구현하기가 까다롭다는 단점이 있다. 이렇듯 완전 동형 암호와 부분 동형 암호는 그 장단점이 선명하고 어느 것이 더 좋은 방법이라고 이야기하기가 어렵다. 본 연구에서는 완전 동형 암호에 대한 내용으로 논의를 진행하고자 한다.

비선형 활성 함수(non-linear activation function)는 딥 러닝의 발전에 가장 큰 기여를 한 기술 중 하나이고, 활발하게 사용되는 비선형 활성 함수의 예로 sigmoid, ReLU를 들 수 있다. 이러한 비선형 함수는 평문 상에서는 연산하기 상당히 쉬운 함수이나, 덧셈과 곱셈만 지원하는 암호문 상에서의 정확한 연산은 불가능하다. 딥 러닝에서 사용되는 주요 연산(합성곱 신경망 연산 등)은 대체로 선형 함수이기 때문에 동형 암호로 구현하는 것이 가능하다. 따라서 비선형 활성 함수를 어떤 방법으로 연산하느냐가 중요한 문제가 되고, 이에 대한 해결을 위해 크게 두 가지 방향의 연구가 진행되어 왔다. 첫째는 암호문 상에서 계산하기 쉬운 새로운 비선형 함수를 제안하는 것이다. 완전 동형 암호에서도 어렵지 않게 계산할 수 있는 $f(x)=x^2$과 같은 단순한 이차, 삼차 다항식을 사용하여 이에 맞는 새로운 딥 러닝 모델을 설계한 연구이다[3, 4, 5]. 둘째는 성능 개선에 유리한 ReLU 함수를 할 수 있는 만큼 최대한 정확하게 근사하여 기존의 딥 러닝 모델에 적용하는 연구이다[6]. ReLU 함수 $f(x)=\max\{x, 0\}$를 동형 암호에 적용하기 위해서, max 함수 또는 이를 위해 필요한 부호 연산(sign function)을 다항식으로 근사하는 연구가 진행되었고[7, 8], [6]은 [7]에서 제안한

minimax polynomial의 합성 기법을 사용하여 CIFAR-10와 ImageNet을 덧셈과 곱셈만으로 높은 정확도로 분류하는 데 성공하였다. 특히 [6]에서 제안한 방법은 ResNet과 같은 실제 성능이 검증된 기존의 딥 러닝의 모델의 구조와 기학습(pre-trained)된 파라미터를 그대로 활용할 수 있다는 장점이 있고, ReLU 함수를 정확하게 근사할수록 이미지 분류의 성능이 정확해진다. 본 연구에서는 ReLU 함수의 정확도에 대한 분석을 보다 자세히 진행하여 실제 암호화된 딥 러닝 모델에서 사용 가능한 다항식이 어떤 것인지, 특히 근사 범위가 어느 정도로 허용되는지에 대한 것을 알아보고자 한다.

Ⅱ. 본론

ReLU 함수를 계산하기 앞서 [7]에서는 아래와 같은 부호 함수

$$sgn(x)=\begin{cases}x/|x| & x\neq 0\\ 0 & x=0\end{cases}$$

를 다항식으로 근사하였다. 구체적으로, 주어진 precision parameter α에 대해, $-1\le x\le 1$의 구간에서

$$|sgn(x)-p_\alpha(x)|\le 2^{-\alpha}$$

를 만족하는 다항식 p_α가 제시되어 있다. 그리고 [6]에서는 이 다항식을 활용하여 $-B\le x\le B$의 구간에서 ReLU 함수 $f(x)=\max\{x,\ 0\}$를 RIGHT

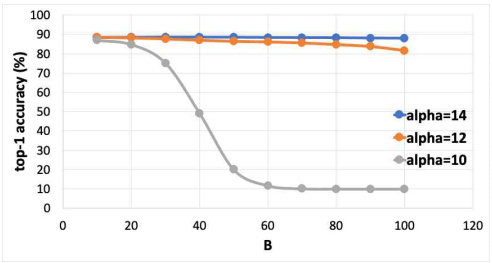
$$r_{\alpha,B}(x)=\frac{1}{2}(x+xp_\alpha(\frac{x}{B}))$$

와 같이 근사하였다. 이 다항식을 사용하면

$$|\ ReLU(x)-r_{\alpha,B}(x)\ |\ \le B(1/2)^{-\alpha}$$

의 오차 관계를 얻게 된다. Precision parameter α에 대해서는 모델의 성능 분석이 진행된 바 있지만[6], 근사 범위에 해당하는 값인 B 역시 ReLU 함수의 오차에 영향을 미치므로 여기에 대한 추가 분석이 필요하다.

본 연구에서는 CIFAR-10을 사용하여 기학습되어 88.36%의 top-1 정확도를 갖는 ResNet-20을 사용하였다. ResNet-20 안의 ReLU 함수를 위에서 언급한 다항식 $r_{\alpha,B}$로 바꾸었을 때 어떤 성능의 변화가 일어나는지에 대한 시뮬레이션을 진행하였다. Precision parameter α는 10, 12, 14를 사용하였고, 근사 범위에 해당하는 값인 B는 10부터 100까지 10을 간격으로 하여 실험하여 top-1 성능의 변화를 관찰하여 이를 [그림 1]에 정리하였다.



〔그림 1〕 B 값에 따른 ResNet-20의 top-1 정확도

[그림 1]에서 볼 수 있듯이 B가 증가할수록 top-1 정확도는 감소하며 이는 식 (1)에서도 유추할 수 있는 사실이다. α가 12와 14일 때는 B가 증가하여도 그렇게 큰 정확도 감소를 보이지 않는다. α가 14일 때는 87.94%까지 성능 저하가 일어났으며, α가 12일 때는 81.50%까지 일어났다. 그런데 α가 10일 때는 성능 저하가 B 값에 상당히 민감하게 일어남을 알 수 있다. B 값이 100까지 늘어나도 α가 12와 14일 때는 어느 정도 유의미한 성능을 얻지만, α가 10일 때는 B 값이 40 정도만 되어도 기학습된 모델의 정확도가 49.05%로 성능에 큰 손실이 있었다. 이는 특정 precision parameter α에 대해서는 B 값 역시 정밀하게 조절해야 올바른 근사 모델을 얻을 수 있다는 것을 의미한다.

B 값이 매우 작으면 정확도는 올라가겠지만, 기학습된 모델의 파라미터들이 어떤 값인가에 따라 B 값을 줄이는 데에 한계가 있다. 파라미터들의 값이 크면 모든 값들이 근사 범위 $-B\le x\le B$에 들어오기가 쉽지 않기 때문이다. 따라서, precision parameter α와 별개로 B에 대한 값 역시 정밀한 조절이 필요한 것으로 요구된다.

Ⅲ. 결론

본 연구에서는 완전 동형 암호 상에서 딥 러닝 모델을 적용하기 위해, ReLU 함수를 다항식으로 근사한 상황에서 근사 범위가 성능에 어떤 영향을 주는지 확인하는 시뮬레이션을 ResNet-20 모델 상에서 진행하였다. 그 결과, 특정 precision parameter α에 대해서는 근사 범위를 정하는 데 필요한 값인 B에 민감하게 성능이 변화하는 것을 확인하였다. 따라서 B 값 역시 기학습된 모델의 성능에 중요한 역할을 하고 있음을 알 수 있었다.

참 고 문 헌

[1] Gentry, C. (2009, May). Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 169-178).

[2] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017, December). Homomorphic encryption for arithmetic of approximate numbers. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 409-437). Springer, Cham.

[3] Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., & Wernsing, J. (2016, June). Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In International conference on machine learning (pp. 201-210). PMLR.

[4] Hesamifard, E., Takabi, H., & Ghasemi, M. (2017). Cryptodl: Deep neural networks over encrypted data. arXiv preprint arXiv:1711.05189.

[5] Chou, E., Beal, J., Levy, D., Yeung, S., Haque, A., & Fei-Fei, L. (2018). Faster cryptonets: Leveraging sparsity for real-world encrypted inference. arXiv preprint arXiv:1811.09953.

[6] Lee, J., Lee, E., Lee, J. W., Kim, Y., Kim, Y. S., & No, J. S. (2021). Precise Approximation of Convolutional NeuralNetworks for Homomorphically Encrypted Data. arXiv preprint arXiv:2105.10879.

[7] Lee, E., Lee, J. W., No, J. S., & Kim, Y. S. (2020). Minimax Approximation of Sign Function by Composite Polynomial for Homomorphic Comparison. IACR Cryptol. ePrint Arch., 2020, 834.

[8] Cheon, J. H., Kim, D., & Kim, D. (2020, December). Efficient homomorphic comparison methods with optimal complexity. In International Conference on the Theory and Application of

Cryptology and Information Security (pp. 221-256). Springer, Cham.