보안 기계학습을 위한 양자-고전 하이브리드 샘플링 프로토콜

송우영 [1], 방정호 [2,*], 이진형 [1,†]
[1] 한양대학교, [2] 고등과학원

*jbang@kias.re.kr, †hyoung@hanyang.ac.kr

Classical-Quantum Hybrid Sampling Protocol for Secure Machine Learning

Wooyeong Song[1], Jeongho Bang[2,6,*], Jinhyoung Lee[1,†]
[1]Hanyang Univ., [2] Korea Institute for Advanced Study

*jbang@kias.re.kr, †hyoung@hanyang.ac.kr

요 약

기계학습에서의 보안에 대한 이슈가 주목받음에 따라, 보안 기계학습을 위한 고전-양자 하이브리드 샘플링 프로토콜을 제안했다. 본 프로토콜을 통해 외부의 적대 학습자 (adversarial learner) 의 존재를 감지하고 적대 학습자의 적법 학습자 (legitimate learner) 와 동등한 수준의 학습을 방지하며, 적법 학습자에게만 학습에 충분한 수의 샘플을 보장할 수 있다. 특히, 보안성이 quantum no-broadcasting theorem 에 기반하기 때문에 학습 샘플이 복제될 수 있는 고전 영역에서는 이러한 형태의 보안을 확보하는 것이 불가능하다.

Ⅰ. 서 론

기계학습이 산업과 일상 전반에서 널리 활용됨에 따라 학습에서의 보안은 속도 향상에 대한 연구와 더불어 오늘날 기계학습 커뮤니티에서의 중요한 이슈로 떠올랐다. 그럼에도 불구하고 기계학습에서의 보안성을 개선하고자 하는 연구인 보안 기계학습 (secure machine learning) 에서의 양자 특성의 활용은 많이 연구되지 않았다 [1-3]. '보안 학습' 이라는 용어는 주로 오직 '적법 학습자 (legitimate learner)' 에게만 학습을 허용하며 허용되지 않은 '적대 학습자 (adversarial learner)' 는 동등한 수준의 학습이 불가능한 학습을 의미한다 [4, 5]. 이러한 적대 학습자의 주 목표는 적법 학습자와 같은 수준의 학습을 하거나, 적법 학습자가 학습을 실패하게 만들거나, 적법 학습자가 목표와 전혀 다른 학습을 하게 만드는 것이다. 자율 주행, 질병 진단 등 인명과 직결되는 여러 분야에서 활용되는 기계학습에서의 보안은 중요한 요소이며, 양자 통신 등에서 보안을 목적으로 활용되는 양자 특성의 기계학습에의 적용은 그 가능성이 유망할 것이다.

Ⅱ. 본론

우리는 학습자가 학습 샘플을 취하는 과정에서 적대 학습자가 존재하는 경우에도 안전하게 샘플링 할 수 있는, 두 적법 학습자 간의 고전-양자 하이브리드 샘플링 프로토콜을 제안했다 [6]. 이 프로토콜을 통해 적법 학습자는 목표로 하는 정확도의 학습을 하기에 충분한 수의 샘플을 가질 수 있으며, 적대 학습자가 도청을 통해 샘플을 빼내더라도 적대 학습자는 적법 학습자 수준의 정확한 학습에 필요한 샘플을 가질 수 없다.

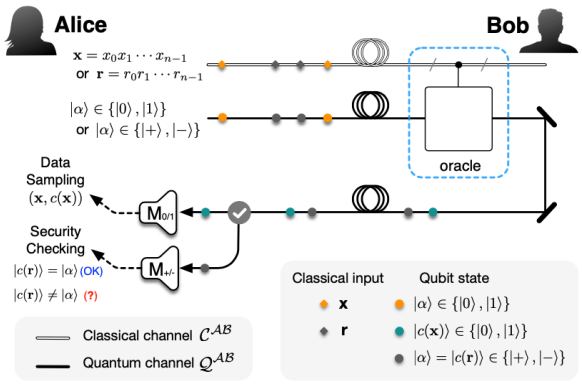

그림 1. 샘플링 프로토콜의 모식도.

Ⅲ. 결론

적대 학습자가 샘플을 중간에서 가로채더라도 적대 학습자는 학습에 충분한 샘플을 가질 수 없고, 프로토콜상 적법 학습자는 적대 학습자의 존재를 알아챌 수 있다. 이러한 성질은 'Quantum no-broadcasting theorem' [7] 에 기반하기 때문에 고전 기계학습에서는 구현이 불가능하다는 특성을 가진다. 또한, 본 모델은 고전-양자 간의 융합 즉, (큰 크기의) 입력 데이터는 고전 정보로 유지한 채 작은 크기의 양자 계로 구성되어, 고전 데이터를 양자 중첩 샘플로 인코딩 하는 과정을 필요로 하지 않기 때문에, 구현이 용이하다는 장점을 가진다. 이에 따라, 오류가 존재하는 중규모의 양자 계를 활용하고자 하는 연구방향인 NISQ 기술에도 적합하다.

ACKNOWLEDGMENT

W.S. and J.B. acknowledge the research project on developing quantum machine learning and quantum algorithm (No. 2019-100) by the ETRI affiliated research institute. W.S. and J.B. acknowledge the financial support of the National Research Foundation of Korea (NRF) grants (2019R1A2C2005504 and NRF-2019M3E4A1079666), funded by the MSIP (Ministri of Science, ICT and Future Planning), Korea government. W.S. and J. L. supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2020-2015-0-00385) supervised by the IITP(Institute for Information & communications Technology Promotion).

참 고 문 헌

[1] J. Bang, S.-W. Lee and H. Jeong, Quantum Information Processing **14**, 3933 (2015).

[2] Y.-B. Sheng and L. Zhou, Science Bulletin **62**, 1025 (2017).

[3] N. Liu and P. Rebentrost, Physical Review A **97**, 042315 (2018).

[4] M. Barreno, B. Nelson, A. Joseph and J. D. Tygar, Machine Learning **81**, 121 (2010).

[5] B. Nelson, S. Rao and J. D. Tygar, J. Mach. Learn. Res. **13**, 1293 (2012).

[6] W. Song, Y. Lim, H. Kwon, G. Adesso, M. Wiesniak, M. Pawlowski, J. Kim and J. Bang, arXiv:1912.10594 (2019)

[7] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa and B. Schumacher, Physical Review Letters **76**, 2818 (1996)