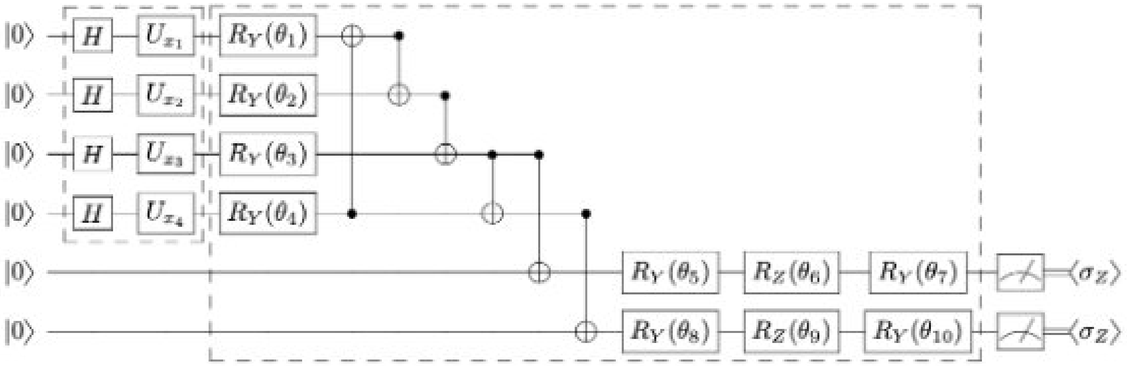
Study tests the potential of two quantum machine learning algorithms for malware classification

May 25 2023, by Ingrid Fadelli



The representation of the QNN used by the researchers. Credit: Barrué & Quertier

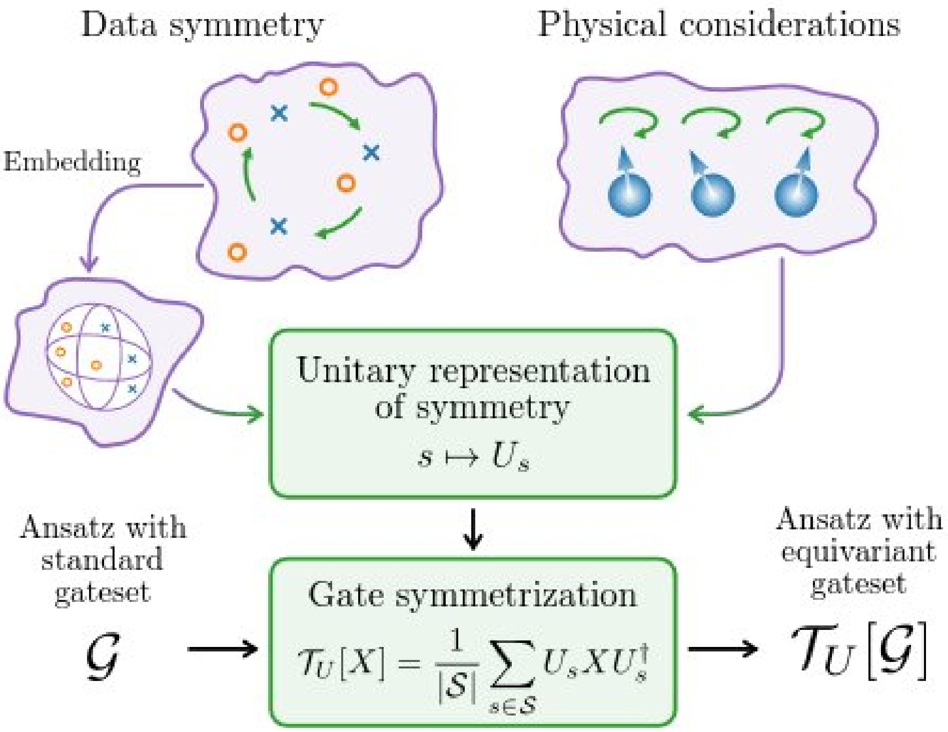
Over the past decades, cyber attackers have become increasingly skilled at compromising systems and circumventing security measures. As a result, detecting and accurately identifying malware is a pressing challenge for many businesses and individuals worldwide.

Cyber-security experts have recently been exploring the potential of machine learning techniques for classifying malware and determining what actions should be taken to eradicate it. While some of these techniques achieved promising results, studies showed that many of them can be fooled or fail to accurately identify malware that they never encountered before.

In the hope of identifying more reliable methods to classify malware, researchers at Orange Innovation Inc. recently carried out a study assessing the potential of the quantum version of machine learning algorithms. Their paper, pre-published on *arXiv*, offers some initial insight into the strengths and limitations of two types of quantum machine learning models, outlining directions that could be explored in future cyber-security research.

"I have been working on using artificial intelligence for malware analysis since 2019," Tony Quertier, co-author of the paper, told Tech Xplore. "With Grégoire Barrué, who started his post-doc in October, we want to explore what quantum technology can bring to this problem. As we both have a mathematical background in two complementary areas, we hope to be able to take advantage of our theoretical knowledge to understand this subject."

Quertier and Barrué believe that quantum machine learning could allow users to extract more information from less data. To test this hypothesis in the context of malware classification, they so far assessed the performance of two different quantum machine learning models, known as QSVM and QNN.


Data symmetry
Physical considerations
Embedding
Unitary representation of symmetry
$s \mapsto U_s$
Ansatz with standard gateset
$\mathcal{G}$
Gate symmetrization
$\mathcal{T}_U[X] = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} U_s X U_s^\dagger$
Ansatz with equivariant gateset
$\mathcal{T}_U[\mathcal{G}]$

Credit: Meyer et al, *PRX Quantum* (2023). DOI: 10.1103/PRXQuantum.4.010328

"The first algorithm we tested is a simple QSVM, an adaptation of the Support Vector Machines algorithm in quantum," Quertier explained. "We then also tested a QNN, a quantum adaptation of a classical neural network. We find the results very encouraging, as we trained them on few data and with, for the moment, two rather simple optimization approaches (SPSB and data reuploading)."

In the initial evaluations they carried out, Quertier and Barrué found that the QSVM algorithm achieved very promising results, outperforming some of the team's classical SVMs for malware classification on several parameters. The QNN, on the other hand, which was only optimized with data reuploading and using a technique known as SPSB, could classify malware with an accuracy of 87%. This is quite good, considering that it was also trained on a limited amount of data.

"Obviously this accuracy is not as good as our classic versions of the algorithm, but our classic versions are trained on 1 million data, while here we only used 1,000 samples," Quertier said. "For a first approach, this is beyond what we expected. For me, the most interesting thing is the ability of quantum machine learning techniques to learn from limited training data. We have become too dependent on having lots of data and computational resources. However, in some domains, it's not so easy to have a lot of data."

The overreaching goal of ongoing research efforts by Quertier and Barrué is to optimize algorithms so that they can efficiently extract a greater amount of information from a limited amount data. In their next studies, they plan to explore the potential of other quantum versions of machine learning algorithms, such as quantum convolutional networks (QCCNs), while also using mathematics to optimize and better analyze available data.

"For example, Lie theory can allow us to identify the number of parameters to achieve overparametrization (when the model has sufficiently many parameters so that the Fisher Information matrix reaches its maximal rank, and thus has maximal capacity) or even identify symmetries in the data and adapt the quantum gates we use," Quertier added. "In October 2023, a Ph.D. thesis will start on this subject, which will be supervised by Daniel Juteau who is a specialist in this [type of] mathematics."

More information: Grégoire Barrué et al, Quantum Machine Learning for Malware Classification, *arXiv* (2023). DOI: 10.48550/arxiv.2305.09674

Citation: Study tests the potential of two quantum machine learning algorithms for malware classification (2023, May 25) retrieved 27 April 2024 from https://techxplore.com/news/2023-05-potential-quantum-machine-algorithms-malware.html