COMMENT

Dequantizing algorithms to understand quantum advantage in machine learning

Ewin Tang

Understanding whether quantum machine learning algorithms present a genuine computational advantage over classical approaches is challenging. Ewin Tang explains how dequantizing algorithms can uncover when there is no quantum speedup and perhaps help explore analogies between quantum and classical linear algebra.

Will quantum computers someday give superpolynomial speedups for machine learning on classical data? Current evidence suggests that certain other tasks, such as factoring, admit paradigm-shifting quantum speedups. One might then hope that, since quantum systems implicitly manipulate exponentially large matrices, one can harness nature's inherent linear algebra processor to manipulate data exponentially faster than it is possible to do with classical computers. Researchers have put forth many proposals to gain a computational advantage with quantum machine learning (QML) algorithms for domains ranging from recommendation systems[1] to topological data analysis[2], but these proposals typically require unique assumptions that make the comparison to classical algorithms difficult. So, our understanding of speedups in this space is much murkier than it might appear at first glance[3].

To shed light on these quantum advantage proposals we developed a new framework for analysing QML algorithms, which can produce formal evidence against an exponential quantum advantage. The idea is to find 'dequantized' versions of QML algorithms, which are fully classical algorithms that, on classical data, perform only polynomially slower than their quantum counterparts. The existence of a dequantized algorithm means that its quantum counterpart cannot give exponential speedups on classical data.

Prior to this dequantizing framework, the only formal evidence for a quantum linear algebra speedup was that certain problems QML can solve are bounded-error quantum polynomial time (BQP)-complete, meaning that if they cannot produce an exponential speedup, then no quantum algorithm can. BQP-completeness is positive evidence for the existence of a speedup, but has only been shown for a handful of QML problems, notably sampling from a solution to sparse linear regression[4]. Dequantized algorithms nicely complement this technique by giving negative evidence of a quantum speedup, provided that one is given input matrices and vectors as lists of entries, referred to here as being given 'classical data'. Researchers have been able to dequantize a wide swathe of QML algorithms, including essentially all known quantum linear algebra algorithms assuming that the input is 'low-rank' in some way, in that it sits approximately in a low-dimensional subspace[5]. This framework has also been used to refine and investigate the quantum speedups for problems coming from quantum chemistry[6].

It seems counterintuitive that classical linear algebra algorithms can perform nearly as well as quantum ones, even on classical data. In some sense, what dequantization shows is that some quantum linear algebra algorithms do not fully exploit 'quantumness', since they can be mimicked classically using sampling procedures. As a simple example, suppose one has two unit vectors ψ and φ of dimension d, and wishes to compute their overlap $|\langle\psi|\varphi\rangle|^2$. There is a quantum algorithm, the swap test, to compute this: prepare the $\log(d)$-qubit quantum states $|\psi\rangle = \sum_{i=1}^{d} \psi_i |i\rangle$ and $|\varphi\rangle = \sum_{i=1}^{d} \varphi_i |i\rangle$, along with one qubit in the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, where $|i\rangle$ are the computational basis vectors. Then apply a controlled SWAP operation between $|\psi\rangle$ and $|\varphi\rangle$, with the additional qubit as the control, and then measure this qubit in the Hadamard basis; the measurement is an estimate of the overlap, and averaging over more runs of this circuit gives an estimate to 0.01 error with only $O(\log(d))$ quantum gates and a constant number of copies of the input states.

Computing such an overlap using classical computers requires $\Omega(d)$ time, so one might naively conclude that the swap test achieves an exponential quantum advantage in the task of computing overlaps. However, there remains the question: given ψ and φ classically, as a list of entries, how can one efficiently prepare the corresponding quantum states that are necessary to run the swap test? State preparation assumptions like these are common in quantum linear algebra, and the typical ways of

University of Washington, Seattle, WA, US.

e-mail: ewint@cs.washington.edu

https://doi.org/10.1038/s42254-022-00511-w

satisfying them (for example, with specialized quantum data structures) also allow classical computers to draw samples from vectors as in a measurement of the states in the computational basis. One can use these samples to estimate the overlap much faster via a Monte Carlo method: namely, to dequantize the swap test, pull one sample, s, from $|\psi\rangle$ (so the probability of sampling an index i is $|\psi_i|^2$), and then compute the estimator φ_s/ψ_s. This estimator has expected value $\langle\psi|\varphi\rangle$ and variance 1, so by averaging over a constant number of runs, one can estimate the overlap to 0.01 error using $O(\log(d))$ classical gates, assuming that the entries of ψ and φ are specified with $O(\log(d))$ bits. The swap test achieves the same dependence on dimension as the dequantized swap test, so it does not give an exponential speedup on classical data.

The general principle of the dequantized algorithm for overlaps extends to other QML algorithms. When input data is given classically, quantum algorithms must perform pre-processing so that input states can be prepared efficiently; so, it follows that for a fair comparison to demonstrate speedup, a classical algorithm should also be allowed the same forms of pre-processing. In particular, it's reasonable to give classical algorithms access to computational basis measurements from whatever vectors and matrices that a quantum algorithm needs to prepare efficiently. Classical algorithms can then exploit these input data measurements to speed up linear algebra to become time-independent of the dimension. Specifically, sketching algorithms explore how to use randomness to perform a dimensionality reduction and 'sketch' a large matrix A down to a constant-sized normalized submatrix of A, which behaves similarly to the full matrix[7]. The computational basis measurements one can produce in a quantum-inspired input model allow for the efficient estimation of matrix products through Monte Carlo methods[8], which can be applied iteratively to produce dequantized algorithms that achieve surprisingly similar bounds to their quantum counterparts.

Dequantization has two main limitations. First, this technique fails catastrophically when applied to data coming from quantum systems[9]: for example, quantum principal component analysis can be dequantized, yet it produces exponential speedups when the classical algorithm only gets access to the input state's measurement data without amplitudes[10]. Dequantized algorithms cannot work without being given an explicit list of amplitudes, suggesting that QML has the best chance of achieving large speedups whenever classical computation cannot get access to this data (which occurs when input states come from quantum circuits and other physical quantum systems).

Second, dequantization does not yet rule out the possibility of large polynomial speedups on classical data, which could still lead to significant performance improvements in practice with sufficiently good quantum computers. This is still an area of active research.

Finally, one may expect that these exponential speedups of dequantized algorithms may improve over existing classical algorithms. However, we are unaware of any settings where dequantized algorithms give advantage. The reasons for this are subtle, and resemble the difficulties in finding applications for QML algorithms. Right now, the main application of dequantization is to demonstrate barriers to quantum advantage, but we hope that the analogies between quantum linear algebra and classical linear algebra will blossom into a fruitful exchange of ideas between the two fields.

1. Kerenidis, I. & Prakash, A. Quantum recommendation systems. In *Proc. 8th Innovations Theor. Comput. Sci. Conf. (ITCS)* 49:1–49:21 (Schloss Dagstuhl, 2017).
2. Lloyd, S., Garnerone, S. & Zanardi, P. Quantum algorithms for topological and geometric analysis of data. *Nat. Commun.* **7**, 10138 (2016).
3. Aaronson, S. Read the fine print. *Nat. Phys.* **11**, 291–293 (2015).
4. Harrow, A. W., Hassidim, A. & Lloyd, S. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.* **103**, 150502 (2009).
5. Chia, N.-H. et al. Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning. In *Proc. 52nd Ann. ACM SIGACT Symp. Theory of Comput. (STOC 2020)* 387–400 (Association for Computing Machinery, 2020).
6. Gharibian, S. & Le Gall, F. Dequantizing the quantum singular value transformation: Hardness and applications to quantum chemistry and the quantum PCP conjecture. In *Proc. 54th Ann. ACM SIGACT Symp. Theory of Comput. (STOC 2022)* 19–32 (Association for Computing Machinery, 2022).
7. Woodruff, D. P. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.* **10**, 1–157 (2014).
8. Drineas, P., Kannan, R. & Mahoney, M. W. Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix. *SIAM J. Comput.* **36**, 158–183 (2006).
9. Cotler, J., Huang, H.-Y. & McClean, J. R. Revisiting dequantization and quantum advantage in learning tasks. Preprint at https://arxiv.org/abs/2112.00811 (2021).
10. Huang, H.-Y. et al. Quantum advantage in learning from experiments. *Science* **376**, 1182–1186 (2022).

Competing interests

The author declares no competing interests.