ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

Enhanced learning for agents in quantum-accessible environments

Vedran Dunjko[1], Jacob M. Taylor[2,3] and Hans J. Briegel[1]

1- Institut für Theoretische Physik, Universität Innsbruck
Technikerstraße 25, A-6020 Innsbruck, Austria

2- Joint Center for Quantum Information and Computer Science,
University of Maryland, College Park, MD 20742 USA

3- Joint Quantum Institute, National Institute of Standards and Technology
Gaithersburg, MD 20899 USA

Abstract. In this paper we provide a broad framework for describing learning agents in general quantum environments. We analyze the types of classically specified environments which allow for quantum enhancements in learning, by contrasting environments to quantum oracles. We show that whether or not quantum improvements are at all possible depends on the internal structure of the quantum environment. If the environments have an appropriate structure, we show that near-generic improvements in learning times are possible in a broad range of scenarios.

Introduction In the last few years there has been an increasing interest in the potential of quantum improvements in aspects of machine learning and artificial intelligence. For example, the theory and algorithms for classification and clustering utilizing quantum mechanics [4, 5, 1, 6, 2, 3], in both supervised and unsupervised settings have been provided. In the setting of reinforcement learning (RL) [7] quantum information processing has been used to reduce space or time complexity of particular learning algorithms [8, 9]. However, for the general setting of RL, where a learning agent and a task environment *interact quantum-mechanically*, no framework or results have been, to our knowledge, presented so far.

Here, we provide the first steps in this direction. We provide a framework for describing quantum learning agents in general task environments. We analyze the types of environments which allow for quantum enhancements, by contrasting environments to quantum oracles. We show that whether or not quantum improvements are at all possible depends on the internal structure of the quantum environment. If the environments has a suitable structure (allows for *oracular access*), improvements in learning times are possible in a broad range of scenarios that we call luck-favoring settings. The results we provide are also particularly relevant for the class of model-based learning agents [10].

Classical and quantum agent-environment interaction The standard turn-based RL paradigm comprises the percept ($\mathcal{S}$) and action ($\mathcal{A}$) sets which specify the possible outputs of the environment, and the agent, respectively. The agent and the environment interact by sequentially exchanging elements from the percept/action sets. A realized interaction up to time step t, between the agent and the environment, that is a sequence $h_t = (s_1, a_2, s_3, s_4, \ldots, s_{t-1}, a_t), s_i \in \mathcal{S}, a_j \in \mathcal{A}$ of alternating percepts and actions is called *the $t-$step history* of interaction. At the t^{th} time-step, and given the elapsed history h_{t-1}, the behavior of the agent at step t is given by the map $M_A^{h_{t-1}}(s \in \mathcal{S}) \in distr(\mathcal{A})$, where $distr(\mathcal{X})$ denotes the set of probability distributions over the set $\mathcal{X}$. The realized agent's action, given history h_{t-1}, is sampled from the distribution $M_A^{h_{t-1}}(s \in \mathcal{S})$. The environment is specified analogously. The basic diagram illustrating agent-environment interaction is given in Fig. 1 (a).

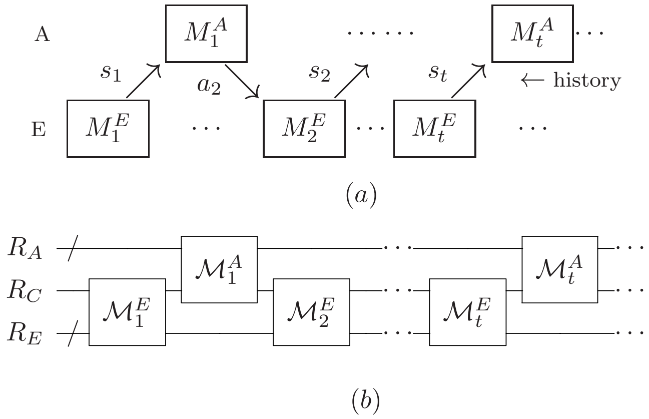

Fig. 1: (a) Classical agent-environment interaction. The maps of the agent and environment may depend on the entire elapsed histories. (b) General interaction between an agent and an environment. There is a unique communication register R_C, visible to both the agent and the environment. The crossed wires represent multiple systems.

The notion of a *reward* $\lambda \in \Lambda$ in RL (specifying whether a performed action, or a sequences thereof were 'correct') can be w.l.g. subsumed into the percept space. All the standard figures of merit regarding the learning performance of an agent are functions of the realized history. They are often convex-linear, enabling statements about average performance. Thus, the interaction history is the fundamental object in RL, which we aim to maintain even in the quantum setting.

In an extension of the above framework to a quantum setting, we promote the percepts/actions to orthogonal states (*percept/action* states) of the percept/action Hilbert spaces $\mathcal{H}_{\mathcal{S}} = \mathrm{span}\{|s_i\rangle\}_i$ and $\mathcal{H}_{\mathcal{A}} = \mathrm{span}\{|a_i\rangle\}_i$. The agent, and the environment, contain internal memory: finite (but arbitrarily sized) *internal registers* R_A and R_E which can store histories, with Hilbert spaces of the form $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}} \cdots$. We jointly the percept/action states, their probabilistic mixtures, and sequences thereof, *classical states.* To model the interaction we define a common *communication register* R_C, with associated Hilbert space $\mathcal{H}_C = \{|x\rangle | x \in \mathcal{S} \cup \mathcal{A}\}$ sufficient to represent both actions and percepts (action and percept spaces are orthogonal)[1]. The agent (environment) is then specified by sequences of completely positive trace preserving (CPTP) maps $\{\mathcal{M}_i^A\}_i$ ($\{\mathcal{M}_i^E\}_i$) acting on the concatenated registers $R_A R_C$ ($R_C R_E$), initialized in a fiducial classical product state. Finally, an agent-environment interaction is defined as the sequential application of the maps, illustrated in Fig. 1 (b).

Classical agents (environments) are those whose maps do not generate non-classical states, given classical states of the internal and the communication registers. Slightly more generally, an agent and environment have a *classical interaction* if, at every stage of the interaction, the joint state of registers $R_A R_C R_E$ is separable w.r.t the three partitions, and the state of R_C, post-selected on any outcome of any separable measurement of $R_A R_E$ is a classical state[2]. To maintain a robust notion of history, we introduce *testers,* systems which monitor the interaction, and record it in its own memory R_T. Testers we consider are sequences of controlled maps of the form

$$U_k^T\left(|x\rangle_{R_C} \otimes |\psi\rangle_{R_T}\right) = |x\rangle_{R_C} \otimes U_k^x |\psi\rangle_{R_T}$$

where $x \in \mathcal{S} \cup \mathcal{A}$, and $\{U_k^x\}_x$ are unitary maps acting on the register R_T, for all steps k. A tested interaction is shown in Fig. 2. If all the maps of the tester copy[3] the classical states we call it a *classical tester.* A *sporadic classical tester* allows periods of untested interaction (i.e. some maps are identities).

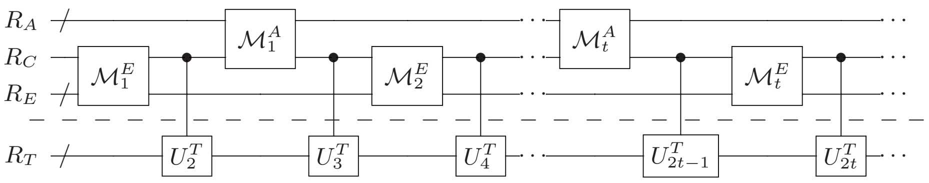

Fig. 2: Tested agent-environment interaction. Note that, in general, each map of the tester U_k^T acts on a fresh subsystem of the register R_T, which is outside the control of the agent and environment. The crossed wires represent multiple systems.

The (generalized) history of interaction is given by the reduced state of the register R_T, and in the case of classical agents, environments and tester, we recover the classical definition of history. With the quantum framework in place we obtain the basic results (Lemmas 1-3 in [11]): 1) an agent-environment interaction is classical iff the state of $R_A R_C R_E$ is the same in the presence of a classical tester; 2) no quantum improvements are possible if the interaction is classical, relative to any tester; 3) no quantum improvements are

[1] A more general definition of an interaction, where in the spirit of robotics and embodied cognitive sciences we separate the interfaces of the agent and the environment, is provided in [11].

[2] Classical interaction still allows that the internal information processing of the agent and environment includes non-classical states, e.g. an internal quantum computer. However, neither the agent or environment are allowed to output non-classical states, or to be entangled to the communication register R_C.

[3] By 'copy' we mean the map $|x\rangle|\epsilon\rangle \rightarrow |x\rangle|x\rangle\ \forall x \in \mathcal{S} \cup \mathcal{A}$, for some fixed state $|\epsilon\rangle$.

possible, relative to the classical tester. Here, by no 'quantum improvement' we mean that for any quantum environment/agent, there exist a classical environment/agent, which generate the same history, given the same tester. These basic results also show no generic quantum improvements are possible just by granting 'quantum access' to a classically specified environment. Specifically, every classical specification of an environment can be realized by a sequence of quantum maps which also simulate classical testing, preventing any quantum improvement relative to any tester by Lemma 3.

Quantum improvements in learning We now focus on categorical examples of (classical) task environments, and consider settings in which learning improvements are possible. Specifically, we focus on epoch-type environments. In such environments, the internal state of the environment is periodically reset: e.g. in chess playing, after a player looses, the board is re-set; similarly, in maze navigating problems, once the walker (agent) has found the goal (exit), it is returned to the initial position, and the task is repeated. Our overall approach is based on contrasting environments to oracles as utilized in quantum algorithms. Although standard environments do not match the specification of oracles (e.g. the actions of the agent are lost to the environment, and not returned), for the class of epoch-type task environments E, we can define *oracular instatiations* E_q of the classically specified environment E, which are unitary. Using these types of instatiations, the agent can perform amplitude amplification[12] in order to obtain a rewarding sequence of actions. Given a classical environment E, which the agent can, at will, access in its classical (E) or oracularized instantiation (E_q), we call controllable environment.

The capacity to find correct sequences of actions faster alone says nothing about the learning capacities of the agent. Nonetheless, an *exploration* stage, i.e. searching, must precede an *exploitation* stage, and the correct interplay of these two phases is a well-studied problem in RL [7, 10][4]. To formalize the intuition that already faster searching may aid in learning, we define *luck-favoring settings*. Roughly speaking, a learning model/agent A and an environment E are luck-favoring, relative to some figure of merit R, if a *lucky* agent (one which by chance alone finds many correct sequences of action during an early stage) outperforms an *unlucky* agent (one which does not) in E, after this early stage and relative to R. We highlight that most benchmarking task environments are luck favoring with most learning models, relative to standard figures of merit (e.g. finite-horizon average reward).

Combining the notions of luck-favoring settings, the capacities of quantum agents to explore faster, and the notion of a sporadic classical tester, we prove the following result (Theorem 1 in [11]):

Main Theorem *(informal) Given a classical learning agent A, and a controllable, classically specified epochal task environment E such that (A, E) are luck-favoring relative to some figure of merit R, there exists a quantum agent*

[4] A related interplay occurs in optimization problems, where a local minimum is often rapidly found, and can be used. However, this opens the possibility that we are missing out on a global minimum.

A^q which outperforms A in R, relative to a chosen sporadic classical tester.

The basic idea is for the quantum agent A^q to use quantum, and untested, access to E_q to obtain instances of winning sequences in quadratically reduced time. Given these sequences, A^q will, internally, and by using no interaction steps, 'train' a simulation of A. Eventually, the simulation will get lucky (call this successfully trained simulation A_{lucky}). Then A^q relinquishes control to A_{lucky}, and forwards percepts and actions between A_{lucky} and the environment, under the classical tester. As the setting is luck-favoring, the main claim follows. For some settings even an exponential separation (in task environment size) can occur, in performance between quantum and classical agents. Exponential separations occur when lucky instances occur exponentially infrequently (e.g. mazes with low connectivity). In this scenario, a quantum agent can find a successful instance with near unit probability while the classical agent still has an exponentially small chance of the same outcome. For time-limited games, this is a relevant exponential improvement.

In the remainder of this work, we show how our approach can be generalized to a wider class of task environments, with minor changes to the definition of the oracular instantiation of the environments. Following this, we consider the problem of oracularization of environments and identify two possibilities in which this can be achieved. The first considers so-called model-based learning agents (MBLA) where the agent constructs an explicit representation of the task environment, and uses it to find optimal responses. In this setting the internal environment is constructed, so our approach provides a generic method for speeding up the internal processing of the agent. The second possibility considers additional options to the agent, including limited access to the internal registers of the environment, using which the agent can manipulate the environment E to behave like the oracular instantiation. While these options may not always be realistic, again in the setting of MBLA they provide a generic method of transforming classically specified environments to oracularized instantiations.

The framework we have established is a quantum generalization of reinforcement learning. This generalization maintains the key notion of history through the concept of a tested interaction. We have presented first results which show quantum improvements in learning in classically specified environments. In the framework we established, other improvements for learning in unknown environments may be possible if different types of testers are considered.

VD and HJB acknowledge the support by the Austrian Science Fund (FWF) through the SFB FoQuS F 4012, and the Templeton World Charity Foundation grant TWCF0078/AB46. JMT acknowledges the support of the National Science Foundation under Grant No. NSF PHY11-25915. JMT thanks HJB and VD for their hospitality during his stays. VD thanks Christopher Portmann and Petros Wallden for useful discussions and comments on the manuscript.

References

[1] Sasaki, M. & Carlini, A. Quantum learning and universal quantum matching machine. *Phys. Rev. A* **66**, 022303 (2002). URL http://link.aps.org/doi/10.1103/PhysRevA.66.022303.

[2] Neven, H., Denchev, V. S., Rose, G. & Macready, W. G. Training a binary classifier with the quantum adiabatic algorithm (2008). arXiv/0811.0416.

[3] Neven, H., Denchev, V. S., Rose, G. & Macready, W. G. QBoost: Large Scale Classifier Training with Adiabatic Quantum Optimization. ACML JMLR Proceedings 25 333–348 (2012).

[4] Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum algorithms for supervised and unsupervised machine learning. *ArXiv:1307.0411* (2013).

[5] Aïmeur, E., Brassard, G. & Gambs, S. Quantum speed-up for unsupervised learning. *Machine Learning* **90**, 261–287 (2013). URL http://dx.doi.org/10.1007/s10994-012-5316-5.

[6] Servedio, R. A.& Gortler, S. J. Equivalences and Separations Between Quantum and Classical Learnability *SIAM J. Comput.* **33(5)**, 1067-1092 (2004).

[7] Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* (MIT Press, Cambridge Massachusetts, 1998), first edn.

[8] Dong, D., Chunlin, C. & Zonghai, C. Quantum Reinforcement Learning. *Advances in Natural Computation* Lecture Notes in Computer Science 3611 686-689 (2005).

[9] Paparo, G. D., Dunjko, V., Makmal, A., Matrin-Delgado, MA., Briegel, H. J. Quantum speedup for active learning agents. *Phys. Rev. X* **4**, 031002 (2014). URL http://journals.aps.org/prx/abstract/10.1103/PhysRevX.4.031002.

[10] Russel, S. J. & Norvig, P. *Artificial intelligence - A modern approach* (Prentice Hall, New Jersey, 2003), second edition edn.

[11] Full technical version of the paper available online at http://arxiv.org/abs/1507.08482.

[12] Brassard, G., Hoyer, P., Mosca, M. and Tapp, A. Quantum Amplitude Amplification and Estimation *arXiv:quant-ph/0005055* (2000).