Deep Learning and Adaptation in e-Learning Systems

L. Georgieva and A. Christev Heriot-Watt University,
Edinburgh, UK, EH14 4AS

***Abstract**: In this paper we study the mechanism and application of deep learning techniques in the context of e-Learning in a multi-agent system. We use an underlying multi-agent system to simulate agents' behaviour and determine the best (perturbed) strategy for reaching equilibrium of the stochastic system. We propose a unifying learning framework to develop and analyse the properties of the equilibrium outcomes for e-Learning multi-agent system, where the agents have clearly defined goals and capabilities. These results potentially have a wider reach to other situations of social interaction often encountered in other domains, for example, economics*

***Key words**: Deep learning, e-learning, multi-agent systems*

INTRODUCTION

Economists, mathematicians and computer scientists have different understanding of formalisms for economical and computational modelling of learning and adaptation, including uncertainty and uncertainty modelling. Prominent formalisms are Q-learning [1], Bayesian Learning [2], Algorithmic Game Theory [3], Neural Networks [4].Successful methods for deep learning often involve artificial neural networks.

Neural networks consist of simple, adaptive processing units, which makes them suitable for modelling and analysis of learning problems in various domains, including e-Learning.

In this paper, we consider and characterize as a unifying learning framework, of which the above algorithms form part. Our framework models a simple scenario where the multi-agents employ reinforcement learning and adaptation, Q-learning and Bayesian learning in a random system with the predefined goal to get the set of answers to a quiz correctly. The formulated framework is applicable to wide range of applications, in which the agents have pre-defined goal, which is useful for measuring the utility of their actions.

In this paper we consider the adaptive application of deep learning methods to e-Learning modelled in a multi-agent system. We show that modelling and analysis of equilibrium state in a e-Learning multi-agent system us a non-trivial problem, especially when the distribution of answers and the invested effort by the agents is considered over time.

Multi-agent systems are a suitable formalism for modelling e-Learning as they allow for the problem of knowledge acquisition to be considered in a distributed environment where agents have their own capabilities and goals. In our scenario, as it is common, the agents are governed by the notion of utility function. In our e-Learning context the utility function refers to how close the agent is to achieving a full set of correct answers in a quiz, given the time, effort, and number of attempts made in the current attempt.

These parameters introduce additional complexity to the multi-agent system, and as a result we show that the analysis of the behaviour of rational e-Learning agents is a complex problem, even when the properties of the domain are abstracted away.

To simplify the considered scenario further, we assume that the e-Learning agents are not competitive, they do not cooperate or exchange information, and that they only learn from the feedback of their own previous attempts. We also assume that the agents have limited communication possibilities and do not exchange information with other agents, but only receive feedback by the multi-agent system. Scenarios, where each of these parameters is defined differently are possible but more complex.

Several taxonomies have been proposed for modelling multi-agent systems. The most commonly used ones consider agent granularity, heterogeneity of agent knowledge, methods for distributing control, and communication possibilities [12].

In our framework, we consider agents that are: (i) benevolent (non-competitive), (ii) learning (whereas the goal stays the same, the agents modify the answers of attempted questions that we consider and they change over time under the assumption of bounded rationality and increasing utility), (iii) acting independently and having one predefined goal (answer all attempted questions in a quiz correctly). In our scenario, there is no communication among agents. Even if such interaction can be viewed as a stimuli, which is potentially increasing the utility of each agent, we abstract away from it in the considered scenario.

To begin, we study how deep learning and adaptation in random systems modelled as neural networks can be used in the scenario of a multi-agent e-Learning system, where the agents with the capabilities defined above interact.

Developing and Implementing a multi-agent system for collaborative e-Learning is a recognized challenge [5]. E-learning systems can be adaptive and represent and maintain information about each user. In such system, the user model is a representation of information about an individual user that is essential for an adaptive system to provide the adaptation effect, i.e., to behave differently for different users. In this context, when the agent (i.e. user) navigates, for example, from one item to another, the system can present the content adaptively. In order to maintain a model of the user interaction, such a system will need to collect the content about the user interaction adaptively as shown in [6].

To begin, we study how deep learning and adaptation in random systems modelled as neural networks can be used in a scenario of a multi-agent e-Learning environment.

MULTI-AGENT SYSTEMS AND LEARNING

Our model uses an underlying multi-agent system. The agents in our framework have initial capabilities to perform a task. While learning, the agents perform repeated trial and error, learning from previous attempts. Thus the agents must be adaptive. The adaptation will be reached by using knowledge-based non-collaborative feedback-based learning techniques, as a way of information acquisition.

Conventionally, we define policy as the core of a reinforcement learning agent in the sense that the behaviour of the agent is determined by the policy. In general, policies may be stochastic and we consider stochastic policies over time in the multi-agent l-Learning system.

A reward function defines the goal in a reinforcement learning problem. We use reinforcement learning as a modelling formalism as it has recently gained ground as a framework where classical techniques from optimal control and dynamic programming with statistical learning and estimation theory can be combined [7].

In this context, this paper attempts as well to use the framework of stochastic optimal control with path integrals to derive a novel approach to reinforcement learning with parametrised policies (governing the behaviour of learning agents) and learning dynamics (affecting the agents' choice of trial and error while learning). In [7] policy improvements that can be transformed into an approximation problem of a path integral which has no open algorithmic parameters other than the exploration noise have been proposed. The approach, described in [7] is founded in value function estimation and stochastic optimal control, which will be useful for our analysis too.

Reinforcement learning is standardly defined as a type of machine learning, concerned with how agents ought to take actions in an environment so as to maximize cumulative reward. In reinforcement learning the agent acts on its environment and receives some evaluation of its action (reinforcement), but is not told of which action is the correct one to achieve its goal. In our scenario, the agent is given immediate feedback on whether the selected answer of a quiz is correct or not, but is given no guidance on how to select the correct answer. This quiz answering scenario is similar to game playing, where the agents know when they will win or lose (we assume that they are given immediate feedback on their quiz answers and win when all are correct), but the agents do not know

how to make each individual answer's choice. Each answer gives the agent an access to a state. Each state has a reward (utility) associated with it (for simplicity we consider a Boolean value: correct versus wrong answer, but numerical utility, i.e. percentage of correct inputs in an answer is also possible).

The agent's task is then to find an optimal policy, mapping states to actions, that maximize long-run measure of the reinforcement.

Our approach uses model based reinforcement learning: our agents learn the correct answers to the quiz and use this to derive the optimal policy. The learning is active: in our model the agents adapt (change their answers to the quiz questions), as a result of the immediate feedback given to them. The agent sees the sequences of state transitions (answers) and the associated rewards (feedback on correct versus incorrect answer with each transition). The agent then updates the utility value of selected answers, given the training sequence. In this model the agent needs to consider what actions to take (for simplification, in this paper limited to selecting an answer to a question) and what the outcomes might be on learning and receiving rewards in the long run.

The advantages of considering reinforcement learning in this scenario, is that the learning algorithm converges fast as it uses information from the environment. The disadvantages are that at each step (after the selection of each answer) the utility function will need to be updated, making the algorithm computationally expensive.

An action (answering a question in the quiz) has two kinds of outcomes: the agent gains reward on the current experience, or the agent' perception on the selected answer has been changed (she has gained an ability to learn).

In our framework, the agent has immediate feedback on whether the selection is immediately good (right versus wrong answer) and feedback on how close her section is to the optimal state (long term good), where all her answers to the quiz questions are correct.

The agent, just like a student taking the quiz can implement one of the two approaches: randomly make answers' selection (computationally expensive) or maximizing the utility using current model estimate (existing knowledge of correct answers).

For optimal selection of answers quickly, the agent can employ various heuristics, for example choosing answers that she has not tried often, avoiding answers that she believes to be obviously incorrect. Enumeration strategies and employing functional optimization (for the utility function) or genetic algorithms are also possible in such scenarios.

FRAMEWORK

We consider the following modelling framework for our scenario: multiple-player, zero-sum stochastic game, defined by a finite set S of states, finite set of actions for the players, a period payoff function, a distribution over the finite set of states and actions and a discount factor.

At every period the system is at some state, the players choose actions, in our example, the action can be for example, selecting an answer of a quiz. The actions are chosen simultaneously and independently. Then the multi-agent system pays a reward to each player (the reward is optimal if the quiz is answered correctly and the payoff increases with a guess closer to the quiz answers for all questions). The game then moves to a new state in the next period, randomized according to the distribution with every new guess.

Players evaluate their infinite stream of payoffs via the discount factor. This statement of the model is a generalization of the single player dynamic programming model which was studied by Blackwell and Bellman [8] [9]. There are famous results which exist via Shapley [10] who proved that every zero-sum stochastic game admits a value, by imitating the familiar single player argument, and others. This has also been used in economics for example asset pricing with a Bellman equation and the contraction mapping

operators. Later work and results showed that non-zero sum discounted stochastic games also admit perfect Markov equilibria.

We study in particular the model with learning of heterogeneous agents with uncertainty. This application is new to the literature as far as we know.

HETEROGENIOUS E-LEARNING MODEL WITH UNCERTAINTY

The early work of [11] contains the key ingredients that have since become central elements used in studies of scenarios with incomplete information. This is similar to the incomplete information that the agent is facing when trying to get the optimal answer.

The environment remains stationary, as no correct answer changes during the e-Learming interaction. The agents re-evaluate their guess and advance selecting the closest or most likely match at every subsequent iteration. Several strands in the literature then evolved around this setting, but a general challenge here has been that multi-period equilibrium models are hard to analyze, even with the aid of numerical methods, especially when there is aggregated uncertainty. The new approach/ learning framework is better suited to address a number of new questions and confront empirical regularities observed in learning agents' behaviour. All agents have identical preferences (achieving the set of correct answers to the quiz in our example), and have the following utility function when initiating the learning process:

$$E_0 (\sum \beta^t U(C_t))$$

The sum here is over t, where at the beginning of the learning process, the value of t is set to 0. U(Ct)) is the constant relative-risk aversion flow of utility from learning, and β is the discount factor.

The labour endowment of each agents is given by an idiosyncratic labour productivity shock z that assumes a finite number of possible values and follows a first order Markov process distribution with transition matrix(z). There is only one asset , a, (the set of correct answers to the quiz) that agents can use to self-insure against risk. This is an exchange economy where each agent receives a random (nondurable) endowment every period (access to their data and attempted answers to the quiz).

In this model, there exists a constant returns to scale production that converts aggregate capital (correct answers) (K) and aggregate labour (number of attempts) (L) into aggregate output (quiz result) (Y). During each period each agent chooses how much to learn and how much of the learning to save for next period by holding risk free assets (the correct answers to the quiz so far). We use recursive notation (a' denotes next period's value).

The multi-agent state variables are denoted by s= (a, z), where a is asset holdings (correct answers) carried into the current period and z is the labour (random number of answers that the agents need to give until they have the full set of correct answers).

The recursive formulation of the problem can be then written as:

$$V(s) = \max \text{ over } C, a' \quad \text{of} \quad U(C) + \beta E_t [V(a', z')| s]$$

Subject to

$$C+ a' = (1+ r) w + zw, \text{ where } C \text{ is non-negative and } a' \geq a.$$

Where r is the rate of return, w is the reward (grade for each agent in our example), and a is a net (natural) reward limit. At every point in time this model economy can be described by a probability distribution of agents over assets (correct answers) a and earnings shocks z (random events when correct answers are guessed).

A stationary equilibrium for this model e is a set answers, involving number of attempts, aggregate correct answers, number of attempted answers, and an invariant distribution of agents over the state variables of the system such that the decision rules solve the Bellman equation of the problem. It is clear that analysing equilibria of the model is not obvious and straightforward even with the application of numerical methods. The learning framework we aim to develop next is designed to solve and enlarge the analysis of the solutions concepts in the context of such general e-Learning model.

REFERENCES

[1] Watkins, J. and Dayan, P. Technical Note: Q-learning, Machine learning, Volume 8, Number 3, pages 279-292, 1996.

[2] Neal, R. Bayesian Learning for Neural Networks, Addison-Wesley, 1996.

[3] Nisan, N. and Roughgarden, T. and Tardos, E. and Vazirani, V. Algorithmic Game Theory, Cambridge University Press, 2007.

[4] Schmidhuber, J. Deep Learning in Neural Networks: An Overview, Neural Networks, pages 85-117, 2014.

[5] Mahdi. H. and Attia, S. IEEE/ACS International Conference on Computer Systems and Applications, 2008.

[6] Brusilovsky, P. and Millán, E. User Models for Adaptive Hypermedia and Adaptive Educational Systems. ETSI Informática University of Malaga, 2008.

[7] Theodorou, E. A. and Buchli, J. and Schaal, S. A generalized path integral control approach to reinforcement learning, Journal of Machine Learning Research, volume 11, number 1, pages 3137--3181, 2010.

[8] Bellman, R. The theory of dynamic programming, Volume 60, number 6, pages 503-515, 2003.

[9] Bellman, R. and Blackwell, D. Some two person games involving bluffing, National academy of sciences, number 35, pages 600-605, 1954.

[10] Shapley, L. Stochastic games, PNAS 39 (10): 1095–1100.

[11] Bewley, T. A difficulty with the optimum quantity of money, Econometrica, volume 51, pages 1485-1504, 1983.

[12] Stone, P. and Veloso, M. Multiagent Systems: A Survey from a Machine Learning Perspective. In Autonomous Robotics volume 8, number 3, 2000.

ABOUT THE AUTHOR

Assoc.Prof. L. Georgieva, PhD and Assoc. Prof. A Christev, PhD, Heriot-Wat University, Edinburgh, UK, EH14 4AS Phone: +44 131 451 8159, E-mail: L.Georgieva@hw.ac.uk.

The paper has been reviewed.