Co-evolutionary learning: lessons for human education?

Alan D. Blair
Dept. of Computer Science
University of Queensland
4072, Australia
blair@cs.uq.edu.au

Abstract

The goal of this paper is to build a bridge from co-evolutionary machine learning and game theory to education policy. A learning system can be modeled as a meta-game between teacher and student, to which we can apply a game theory analysis. In some cases, learning may be stifled by various forms of *collusion*, which appear as suboptimal equilibria in the *meta-game of learning*. Some recent results in co-evolutionary machine learning suggest that these opportunities for collusion may be avoided if certain features are incorporated into the learning environment, raising important issues for the future of education paradigms and policy.

1. Introduction

All education systems will face enormous challenges and opportunities in the years ahead. The pressure to deregulate and privatise will likely lead to a major restructuring, as publicly funded schools and universities face strong competition from an increasing number of private alternatives. The information revolution will make new means of rapid communication available to both teachers and students. But how can such changes be managed without precipitating a breakdown in the system or gradual erosion of standards?

When it comes to the reward and promotion of teachers, effective learning usually takes a back seat to other factors such as seniority (at the high school level) or research output (at the university level). When teaching is taken into account, it is generally measured in a way that is based not on the effectiveness with which the students have actually been taught, but rather (which is not the same thing) the extent to which the students themselves are satisfied with their overall educational experience. The fact that students and teachers are effectively evaluating *each other* can create opportunities for *collusion* between teacher and student. For example, students sometimes put pressure on their professors to make courses easier, or to narrow the scope of the material. A professor who gives way to such pressure is rewarded with positive teaching evaluations, and a reduction in time spent preparing lessons and listening to student grievances, thus allowing them to publish more papers and earn promotion faster. If funding of departments is dependent on the number of students they attract, it can create an incentive for them to water-down their courses, in order to attract more students. These kinds of collusion can, over time, erode standards substantially.

In (Pollack & Blair, 1998) a possible connection was pointed out between this kind of collusion in educational settings and certain phenomena which have recently been observed in studies of co-evolutionary machine learning. The present work is aimed at fleshing out some of the details of this connection. While ideas from education have been used extensively in the past for the design of machine learning systems, we put forward the reciprocal view that the results of machine learning experiments may also help us to better understand certain aspects of education systems. Section 2 introduces the meta-game of learning analysis in education; Section 3 describes some recent results in co-evolutionary machine learning; Section 4 discusses possible implications of these results for education policy.

2. The Meta-Game of Learning in Education

An education system is a complicated entity with many players including teachers, students, administrators, parents and funding bodies – each with different goals and different reward systems. The complex interactions between these various participants could be modeled using computer simulations and game theory analysis. Hopefully, such *Meta-Game of Learning* (MGL) models (Pollack & Blair, 1998) will help us to understand how learning can be enhanced and opportunities for collusion avoided, and predict the likely consequences of proposed changes in education policy.

As an example, consider a model of a learning institution in which n teachers $\{T_i\}_{1\leq i\leq n}$ are responsible for teaching groups of students. Assume that a new group of students enters in each time period and is taught by a different teacher in each of m time periods prior to graduation (for simplicity, we take $m = n$). We further assume that students will undergo some form of *evaluation* (assignments, examinations, etc.) testing them on the material they have been taught.

Decisions must be made about how much material is taught by each teacher, how the students are to be evaluated, and what grades are assigned to them. Let us assume that the size of the course taught by T_i can be measured as an integer s_i representing the amount of material over which the typical student has demonstrable competence at the end of the course. We may think of s_i being determined by the *curriculum* assigned for the course, the quality and scope of the *teaching*, and the *standard* of competence that students are required to demonstrate in the evaluation.

This model is meant to cover *incremental* changes rather than major reorganisations. If one teacher makes a slight increase or decrease in the material covered by his or her course, then teachers of follow-on courses can easily accommodate this change by adding or dropping the appropriate material from their own courses, as long as the change is not too large. Specifically, we assume that s_i may change by δ_i in each time period, where $\delta_i \in \{-1, 0, +1\}$. The three *strategies* $+1$, -1 and 0 may be thought of as raising or lowering the standard of a course, or leaving it unchanged, respectively.

How then can we define an appropriate 'payoff' for each of these strategies? First we have to take account of the *effort* required to teach a course, which we assume to be a linear function of the material $(s_i + \delta_i)$ it contains:

$$\text{effort} = E_0 + \eta(s_i + \delta_i)$$

where E_0 and η are constants. This term is meant to reflect:

(i) time taken to prepare lectures, assignments and other teaching materials,
(ii) time spent answering students' questions and responding to their complaints,
(iii) the stress of dealing with students' requests for higher grades (which can be alleviated by lowering the standards or truncating the curriculum).

We divide the *benefits* of teaching into two terms: firstly, there is the benefit that a teacher derives *individually* from the teaching of his/her own course, in the form of respect from peers and admiration from students. This we model as:

$$\text{individual benefit} = A_0 + \alpha(s_i + \delta_i)$$

Second, there is the benefit that is shared by all teachers *collectively*, due to the relative ease of teaching students who enter classes better prepared, and the prestige accrued to the institution if students appear well-educated upon graduation.[1] We model this as

$$\text{collective benefit} = B_0 + \beta \sum_{j=1}^{n} (s_j + \delta_j)$$

Putting these terms together, we arrive at the following payoff function:[2]

$$\text{payoff}(T_i) = -[E_0 + \eta(s_i + \delta_i)] + [A_0 + \alpha(s_i + \delta_i)] + [B_0 + \beta \sum_{j=1}^{n} (s_j + \delta_j)]$$

If we assume that each teacher T_i has sole authority to choose δ_i in each time period, then T_i's best strategy is:

$$\delta_i = \left\{ \begin{array}{rl} -1, & \text{if} \quad \alpha + \beta < \eta \\ 0, & \text{if} \quad \alpha + \beta = \eta \\ +1, & \text{if} \quad \alpha + \beta > \eta \end{array} \right.$$

We would argue that society has recently been witnessing a gradual shift towards the region where $\alpha + \beta < \eta$. A number of factors have contributed to this shift. Historically, both students and teachers typically remained in the same geographical region for many years, thus making teachers accountable to each other, as well as employers, parents, former students and the wider community. Moreover, classes were generally taught according to a well-established curriculum, so it was easy for other teachers to notice if students knew more or less than they were "supposed to know" after attending a particular course (which appears as an increase of α in the above equation). In contrast, today's courses are restructured frequently in response to changing technology, community attitudes or educational fads. Academics move with greater rapidity from one place to another, while students typically drift away to another community once their education is complete.

These factors have created a situation in which accountability is reduced because no-one is very sure about who is supposed to know what and at which stage (causing a decrease in α). The mobility of academics also makes them less concerned for the infrastructure and prestige of their current institution (causing a decrease in β).

The current trend towards 'distance education' will likely accelerate both these effects, as face-to-face interaction diminishes and learning becomes more and more anonymous with respect to both teacher and institution. Students in the future may even play a greater role in structuring their own education (see Section 4).

Returning to our game theory analysis, we will assume from now on that $n > 1$ and $\beta < \eta - \alpha < n\beta$. Therefore the optimal strategy for each teacher is $\delta_i = -1$. This may

[1] There should really be a time-delay associated with the prestige payoff reflecting the time taken for students to graduate, but we do not explicitly build this into our model.

[2] Note: η, α and β are the important parameters; E_0, A_0 and B_0 play no role in the game theory analysis.

be recognised as a classic *prisoner's dilemma* (Axelrod, 1984) or *tragedy of the commons* (Hardin, 1968) in which, by making rational decisions independently, all teachers end up suffering a loss of $(n\beta + \alpha - \eta)$ in each time period, when they could instead stand to *gain* the corresponding amount if they co-operated in all choosing $\delta = +1$.

3. Co-Evolutionary Machine Learning

The goal of Machine Learning is to design software systems which can learn, from interacting with their environment, information that will help them to perform particular tasks better. Such systems, if successful, will automatically adapt to new environments without re-programming, thus saving humans the trouble of designing all the relevant features by hand.

The success of a machine learning system depends very much on the learning environment in which it is placed. After it has extracted all the accessible information from its original environment, it may need to be put in a new environment in order to progress. "Curricular" or "staged" learning (Langley, 1995) occurs when a learner is placed into a pre-designed series of environments one after the other, as it progresses. However, designing an appropriate series of environments may be very difficult. This difficulty would be avoided if there were some way for the learner and its environment to *co-evolve* with each other, so that the one would always be appropriate for the other (Axelrod, 1984).

Strategic games provide a good opportunity to study this kind of co-evolutionary learning. In theory, several machine learning systems trying to master a competitive game could all learn to improve their strategies simultaneously by playing each other and observing the outcomes – as each one improved, it would provide a slightly more challenging opponent for the others, fuelling a continuing spiral of advancement (in the MGL framework, each player would act as a teacher for its opponents). While this idea has been around since the early days of Artificial Intelligence, interestingly some applications of it have been very successful while others have run into serious difficulties. These difficulties can generally be put down to various kinds of *collusion* between teacher and student which give rise to suboptimal equilibria in the MGL. One example (e.g. in chess or tic-tac-toe) is where the student and teacher *draw* each other, or take turns 'throwing' alternate games. Another is a *narrowing of scope* in which the players keep playing the same kinds of games over and over, only exploring some narrow portion of the strategy space and missing out on key regions where they would then be vulnerable to humans or other players.

However, there have been a few notable cases in which these problems have apparently been avoided. One such instance came to light when Tesauro (1992) compared two different methods for training neural networks to play the game of Backgammon. The first network was trained on a large database of hand-crafted positions, with corresponding moves chosen by a human expert; the second network was trained by having it play against itself thousands of times and using the outcome of each game to make a small adjustment in its strategy according to the *temporal difference* or TD-learning algorithm (Sutton, 1988). Surprisingly, the network trained by self-play, though it initially played a poor (essentially random) game, eventually surpassed the network trained on the expert database, and a later version called TD-Gammon (incorporating some additional hand-crafted features) achieved world master level play (Tesauro, 1995).

A second example is provided by the *Evolving Virtual Creatures* (EVC) domain. In a game devised by Karl Sims (1995) two virtual creatures compete in a world with simulated physics for control of a cube initially placed between them. In each round of competition, all creatures from one species played against the champion of the other species from the previous round. Over several generations, competing species were observed to leap-frog each other in evolutionary arms races, as they each discovered methods for reaching the cube and then further evolved strategies to counter the opponent's behaviour. Some creatures pushed their opponent away from the cube, some moved the cube away from its initial location and then followed it, while others simply covered up the cube to block the opponent's access.

While the exact reasons for these successes were unclear at first, our hypothesis is that these two domains have special attributes which help to prevent sub-optimal equilibria in the MGL (Pollack & Blair, 1998). Although further work is needed to gain a better understanding of these issues, we can say at this stage that the following features seem to play an important role:

Well-defined evaluation: Backgammon always ends in a win or a loss rather than a draw, thus preventing collusion by repeated draws. In the EVC domain, victory is clearly defined in terms of proximity to the cube at the end of the simulation, and constrained by the (simulated) laws of physics.[3] In biological ecosystems (where the term *co-evolution* originates) success is defined in a clear-cut fashion by survival and reproduction.

Broad spectrum of opportunity: in both these domains, the learner has available to it, at any given time, a number of avenues for improvement. In backgammon, there are many aspects of the game which can be developed independently (e.g. blocking, racing, back-game strategy, etc.). In addition, the dice rolls sometimes allow a weak player to score a victory over a strong one – an experience from which both can learn. Virtual creatures, like their biological counterparts, can improve by developing a slightly longer arm, slightly better sensors, or becoming slightly faster, etc. This is in contrast to some other domains, where learning can only proceed along a set path (learn A, then B, then C, etc. in a pre-determined order).

4. Discussion

The goal of this paper is to suggest some links between co-evolutionary learning and education, in the hope of stimulating further discussion between the two fields. Game domains providing a broad spectrum of opportunity, such as those described in Section 3, may perhaps be compared with open-ended or *constructionist* learning environments, in which students are able to explore ideas for themselves without having to stick to a fixed curriculum (Papert, 1993) and which provide students at all levels of ability with an opportunity to learn. On the other hand, the need for a well defined evaluation in co-evolutionary learning suggests that students, in addition to a rich environment, also need an incentive to explore, and a way of monitoring their own progress. One of Tesauro's key findings was that a neural network trained by co-evolutionary learning played better backgammon than a similar network trained on a database of 'expert preferences'. It would be interesting to see whether this result can be related to the notion of 'situated learning' in education.

[3] Indeed, Sims (1995) reported that creatures were quick to exploit early bugs in his simulation which allowed for non-conservation of energy and momentum.

As noted in Section 2, the student of the future might be expected to act with a great deal of autonomy – auditing pre-recorded lectures or multi-media presentations, working with interactive learning environments, doing assignments either individually or collectively, and consulting the (human) teacher from time to time for general guidance and resolution of particular difficulties. Ideas from co-evolutionary machine learning may prove very helpful in the design of such learning paradigms. In particular, they should encompass a broad spectrum of opportunity, so that students will not get bored with the material or stuck on a particular point at a time when they have no immediate access to a teacher. One approach would be to develop software agents which 'co-evolve' with their human trainees, continually adapting to the needs and interests of individual students (Sklar, Blair & Pollack, 1998).

5. Conclusion

In this paper we have attempted to build a bridge from co-evolutionary machine learning and game theory to education policy. The education system is a complex and highly non-linear entity, and radical changes currently under discussion or already in progress would make it even more complex. Computational simulations and theoretical analyses based on the meta-game of learning framework may provide a better understanding of the likely consequences of these changes, and play an important role in guiding the design of educational paradigms and the future direction of education policy.

Acknowledgements

Many of these ideas were developed in discussions with Jordan Pollack and the DEMO Group at Brandeis University. Thanks to Elizabeth Sklar, Janet Wiles and Uri Wilensky for helpful comments and suggestions. This work was funded by a University of Queensland Postdoctoral Fellowship.

References

Axelrod, R. 1984. *The Evolution of Co-operation*, Basic Books, New York.

Hardin, G. 1968. The tragedy of the commons, *Science* **162**, 1243–1248.

Langley, P. 1995. Order Effects in Incremental Learning, in: *Learning in Humans and Machines: Towards an Interdisciplinary Learning Science*, P. Reiman & H. Spada, eds., Pergamon.

Papert, S. 1993. *The Children's Machine: Rethinking School in the Age of the Computer*, Basic Books.

Pollack, J.B. & A.D. Blair, 1998. Co-evolution in the successful learning of backgammon strategy, *Machine Learning* (to appear).

Sims, K. 1995. Evolving 3D morphology and behavior by competition, Proceedings of the Fourth International Conference on Artificial Life, MIT Press, 28–39.

Sklar, E., A.D. Blair & J.B. Pollack, 1998. Co-evolutionary learning: machines and humans schooling together, in: G. Ayala, ed., *Proceedings of Workshop on Current Trends and Applications of Artificial Intelligence in Education*, ITESM, Mexico, 1998, 98-105.

Sutton, R. 1988. Learning to predict by the method of temporal differences, *Machine Learning* **3**, 9–44.

Tesauro, G. 1992. Practical issues in temporal difference learning, *Machine Learning* **8**, 257–277.

Tesauro, G. 1995. Temporal difference learning and TD-Gammon, *Communications of the ACM* **38**(3), 58–68.