Lecture 8 - Co-evolutionary Learning

Yuhui Shi

CSE, SUSTech

Review of the Last Lecture

- Population diversity, Niching, Speciation
- Co-evolution

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

Outline of This Lecture

- **Introduction**
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

Evolutionary Learning

Very straightforward conceptually!

1. Initialise population, X(t = 1).
2. Evaluate fitness of each population member.
3. Select parents from X(t) based on fitness.
4. Generate offspring from parents to obtain X(t + 1).
5. Repeat steps (2-4) until some termination criteria are met.

Two Approaches to Evolutionary Learning

Things might get a little trickier.

- Michigan Approach: Holland-style learning classifier systems (LCS), where each individual is a rule. The whole population is a complete (learning) system.
- Pittsburgh Approach: Each individual is a complete system. This lecture deals only with the Pittsburgh-style evolutionary learning since it is more widely used.

Current Practice in Evolutionary Learning

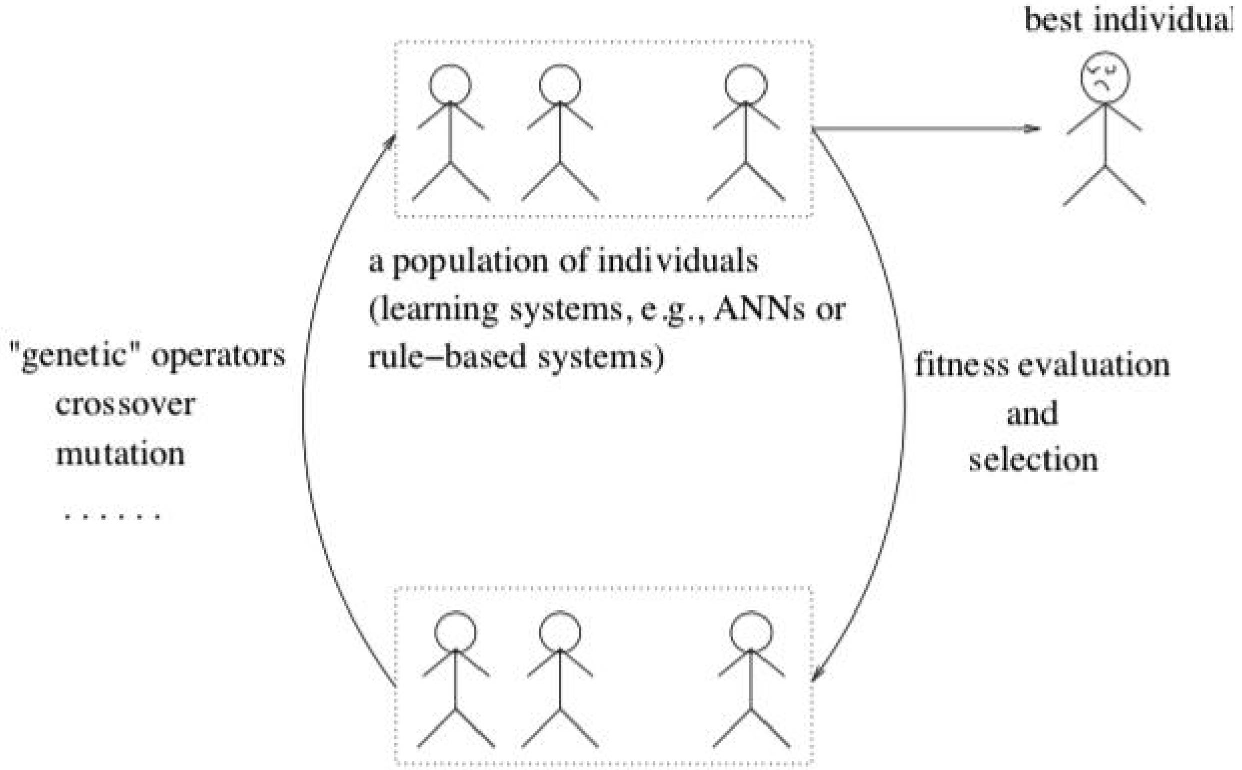


Figure 1: A general framework for Pittsburgh style evolutionary learning.

Fitness Evaluation

1. Based on the training error.
2. Based on the training error and complexity (regularisation), e.g.,

$$\frac{1}{fitness} \propto error + \alpha \cdot complexity.$$

What If No Error Function Is Available

- Or, we don't know how to obtain the fitness function required to evaluate the fitness of a population member, e.g., if we want to evolve game-playing strategies.
- In other words, the exact teacher/target information is unavailable.

Well. . . We have Co-evolutionary Learning

1. Initialise population, X(t = 1).
2. Evaluate fitness through **interactions** between population members.
3. Select parents from X(t) based on fitness.
4. Generate offspring from parents to obtain X(t + 1).
5. Repeat steps (2-4) until some termination criteria are met.

Co-evolutionary Learning In One Slide

1. There has been a HUGE body of literature on co-evolution, especially since Hillis's seminal work in 1991. If we google “Co-evolutionary Learning”, we would get more than 200k hits.
2. Many issues have been raised and discussed: robustness, cycles, mediocre stable states, . . .
3. It is a great idea, but sometimes it just does not do what you hope it would do - frustrating!
4. There is only a small body of literature on theoretical aspects of co-evolution. We need more vigorous theories to move the research forward.

Outline of This Lecture

- Introduction
- **Strategy Games**
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

Iterated Prisoner's Dilemma

	Cooperate	Defect
Cooperate	R R	T S
Defect	S T	P P

Figure 2: Figure 1 of [9]. The payoff matrix for the 2-player prisoner's dilemma (2PD) game. The values S, P, R, T must satisfy T > R > P > S and R > (S + T)/2.

In the 2-player iterated prisoner's dilemma (2IPD) game,

- the above interaction is repeated many times, and
- both players can remember previous outcomes.

Sometimes Two Players Are Not Enough

- Davis et al. [4] commented that

  "The N-player case (NIPD) has greater generality and applicability to real-life situations. In addition to the problems of energy conservation, ecology, and over population, many other real-life problems can be represented by the NIPD paradigm."

- Colman, and Glance and Huberman [2, 5, 6] have also indicated that the NIPD is "qualitatively different" from the 2IPD and that ". . . certain strategies that work well for individuals in the Prisoner's Dilemma fail in large groups".

The NIPD Game

The NIPD game can be defined by the following three properties:

- each player faces two choices between cooperation (C) and defection (D);
- the D option is dominant for each player, i.e., each is better off choosing D than C no matter how many of the other players choose C;
- the dominant D strategies intersect in a deficient equilibrium. In particular, the outcome if all players choose their non-dominant C strategies is preferable from every player’s point of view to the one in which everyone chooses D, but no one is motivated to deviate unilaterally from D.

The NIPD Game - Payoff Matrix

		Number of cooperators among the remaining $n-1$ players				
		0	1	2		$n-1$
player A	C	C_0	C_1	C_2	$\cdots$	C_{n-1}
	D	D_0	D_1	D_2	$\cdots$	D_{n-1}

The payoff matrix of the n-player Prisoner's Dilemma game, where the following conditions must be satisfied:

1. $D_i > C_i$ for $0 \le i \le n - 1$;
2. $D_{i+1} > D_i$ and $C_{i+1} > C_i$ for $0 \le i < n - 1$;
3. $C_i > (D_i + C_{i-1})/2$ for $0 < i \le n - 1$.

The payoff matrix is symmetric for each player.

The NIPD Game - An Example

Number of cooperators among the remaining $n-1$ players

player A	0	1	2		$n-1$
C	0	2	4	$\cdots$	$2(n-1)$
D	1	3	5	$\cdots$	$2(n-1)+1$

Why NIPD Games?

“The NIPD corresponds to a truly remarkable range of real-world social problems, but a few simple examples will suffice.”

- A. Colman, Game Theory and Experimental Games, Pergamon Press, 1982. (pp.157–159) Spring

The Diner’s Dilemma

In game theory, the unscrupulous diner's dilemma (or just diner's dilemma) is an **n-player prisoner's dilemma**. The situation imagined is that several people go out to eat, and before ordering, they agree to split the cost equally between them. Each diner must now choose whether to order the costly or cheap dish. It is presupposed that the costlier dish is better than the cheaper, but not by enough to warrant paying the difference when eating alone. Each diner reasons that, by ordering the costlier dish, the extra cost to their own bill will be small, and thus the better dinner is worth the money. However, all diners having reasoned thus, they each end up paying for the costlier dish, which by assumption, is worse than had they each ordered the cheaper.

Question

What would you do as a rational individual?

Conservation of Scarce Resources

E.g., the energy crisis, the drought in Britain in the summer of 1976, etc.

“It is evidently in each individual's rational self-interest to ignore the call for restraint irrespective of the choices of the others. But - and this is the rub - if everyone pursues individual rationality in this way, they are worse off than if everyone is motivated by collective rationality and exercises restraint. If everyone tries to be a 'free ride' then no-one gets a ride at all.”

The Tragedy of the Commons

First discussed by Hardin in 1968.

- There are six farmers with one cow each weighted at 1000 lb.
- They share a common pasture which can only sustain six cows.
- Each additional cow will reduce the weight of every cow by 100 lb.
- Each farmer is always better off to have an additional cow regardless of the choices of other farmers.
- If they all have one additional cow, they end up with two 400 lb. cows.

Such tragedies are not uncommon in our highly competitive world. It is suggested that "the impoverishment of small farmers in England during the period of the enclosures in the eighteenth century may have been exacerbated by this phenomenon."

Outline of This Lecture

- Introduction
- Strategy Games
- **Co-evolutionary Learning of Game-playing Strategies**
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

Genotypical Representation of 2IPD Strategies

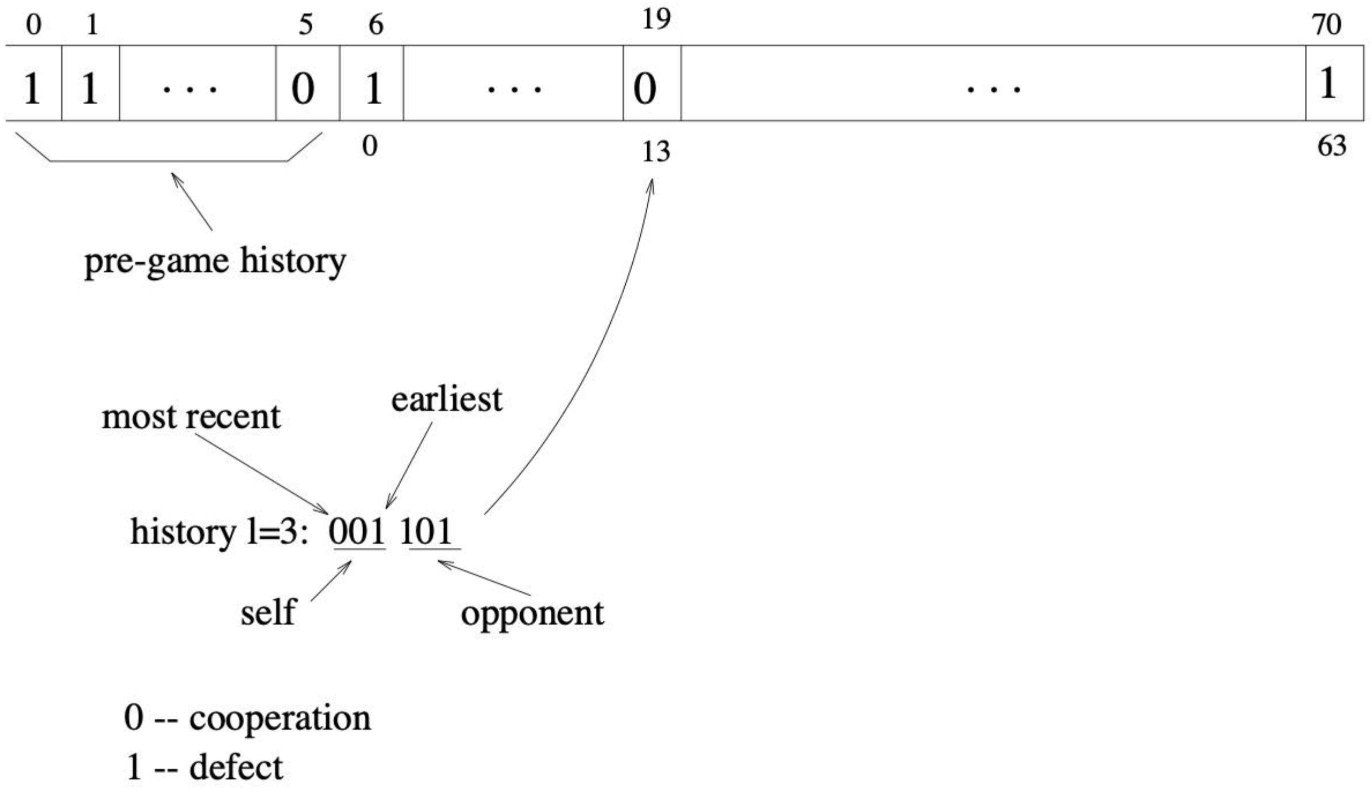


Figure 3: Encoding of strategies, assuming history (memory) length 3.

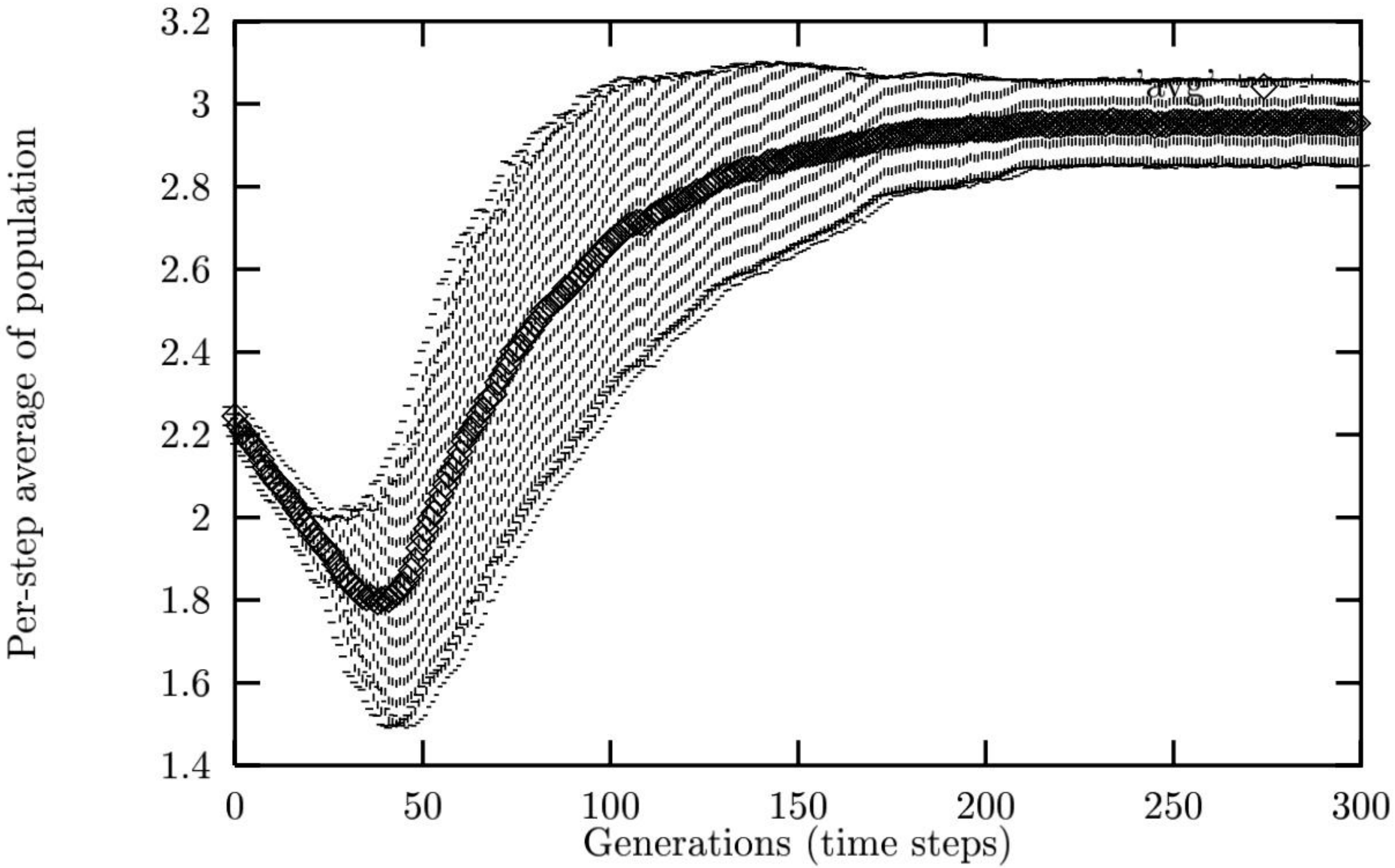


Figure 4: Results from 30 runs of the 2IPD with history (memory) length 3.

Why Co-evolution

1. Co-evolution occurs frequently in Nature.
2. Co-evolution does not require any fixed fitness function.
3. Co-evolution provides a nice approach to learn without expert knowledge and a teacher.

Discussion on the 2IPD Game

1. Cooperative behaviours can be evolved from a population of random strategies without specifying an explicit and fixed fitness function.
2. A cooperative population can be vulnerable to defective intruders.
3. Co-evolutionary learning may not be able to produce strategies which generalise well against unseen opponents.

A Team Approach to Evolving Strategies

1. Evolving a strategy that is capable of dealing with all kinds of opponents can be hard.
2. Maybe we can use a team of individuals as a strategy so that each individual is only dealing a one or two types of opponents. Such individuals should be quite easy to evolve.
3. But how to evolve a team of individuals who are complement of each other?
4. Artificial speciation appears to be an effective technique to achieve that.

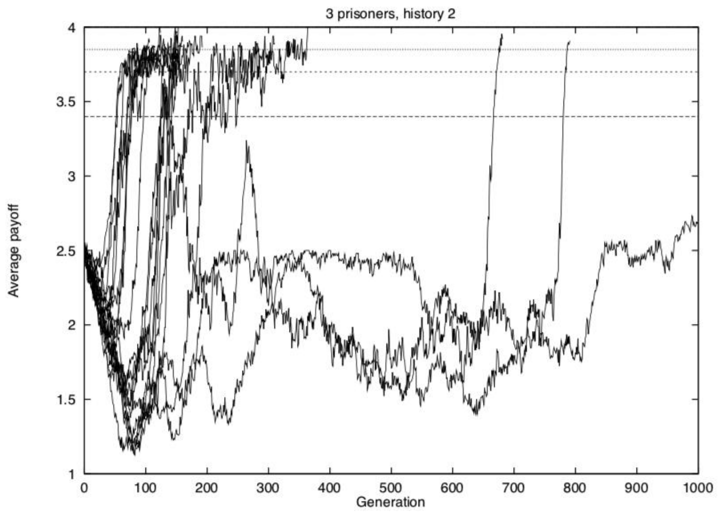
Automatic Modularisation of Learning Systems

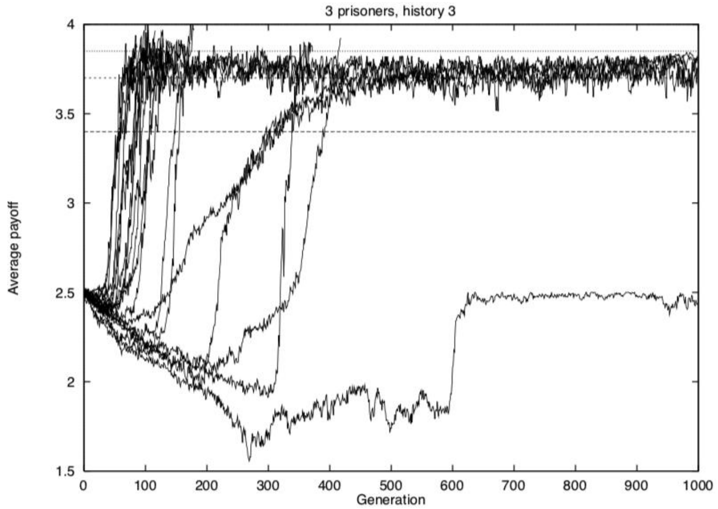
1. A genetic algorithm can be used to evolve strategies for playing the 2IPD.
2. Implicit fitness sharing can be used to form different species (specialists).
3. A gating algorithm can be used to decide which species should respond to an unknown opponent.

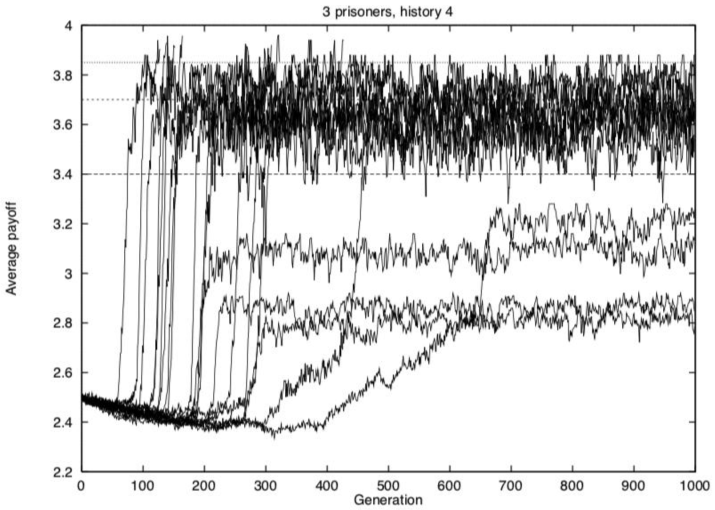
NIPD Game: Some Interesting Questions

1. What is the impact of history length (memory) on the evolution?
2. What is the impact of group size on the evolution?

3IPD with History Length 2, 3 and 4







Discussions on the NIPD Game

1. Memory helps learning. Long history lengths encourage cooperation.
2. Large groups make the evolution of cooperative strategies more difficult. (Not shown in the lecture. Details are in the paper.)

However,

1. We rarely commit ourselves to one way or the other. We prefer to be somewhere in the middle.
2. We are more likely to have short games with short memory.
3. We may not be able to interact with everyone although we know them by names (reputation).

Multiple Levels of Cooperation

- The payoff to player A is given by:

$$p_A = 2.5 - 0.5c_A + 2c_B, (-1 \leq c_A, c_B \leq 1) \quad (1)$$

- where c_A and c_B are the cooperation levels of the two players, which are discretised into four choices of cooperation.

	-1	$-\frac{1}{3}$	$+\frac{1}{3}$	$+1$
-1	1	$2\frac{1}{3}$	$3\frac{2}{3}$	5
$-\frac{1}{3}$	$\frac{2}{3}$	2	$3\frac{1}{3}$	$4\frac{2}{3}$
$+\frac{1}{3}$	$\frac{1}{3}$	$1\frac{2}{3}$	3	$4\frac{1}{3}$
$+1$	0	$1\frac{1}{3}$	$2\frac{2}{3}$	4

Figure 6: The payoff matrix for the 2-player prisoner's dilemma game, with four choices of cooperation.

The Key Question

Question

What is the impact of multiple levels of cooperation on the evolution of cooperation?

Remarks

1. IPD games are useful models of real-life situations.
2. NIPD games are more realistic and general than the 2IPD one.
3. Multiple levels of cooperation discourage cooperation.
4. Co-evolutionary learning is an effective learning technique.

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- **Theoretical Framework of Generalisation in Co-evolutionary Learning**
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

A Simple Research Question

- If I invent a wonderful co-evolutionary learning algorithm and use it to co-evolve a really intelligent game-playing strategy (e.g., for chess, car racing, iterated prisoner's dilemma, or others), how do I know it would perform well **against a new opponent** that it has never seen before?
- Can we say anything at all about the ability (performance) of our co-evolved solutions in a new and unseen environment?
- Sounds like generalisation in machine learning.

Generalisation? We Know That!

1. There have been various discussions about robustness of co-evolved solutions.
2. However, we still do not have any quantitative analysis of generalisation performance, e.g., an absolute quality measure for co-evolved solutions.
3. It is still very hard to compare performance between different co-evolutionary learning algorithms (when applied to a problem).

An Early Attempt

An empirical approach to estimate generalisation of co-evolved solutions:

1. Sample a large number of random test strategies.
2. Co-evolved strategies compete against these test strategies.
3. Generalisation is taken to be the average performance against these test strategies.
4. Note that the number of test strategies should be significantly larger than what co-evolution can search (so that the vast majority of the test strategies are unseen by co-evolved strategies).

Paul J Darwen and Xin Yao. “On evolving robust strategies for iterated prisoner‘s dilemma”. In: Progress in Evolutionary Computation. Springer, 1993, pp. 276–292

The Need for a General Theoretical Framework

1. Although the empirical estimation can give us some information, it is unknown how accurate the estimate is to the true value.
2. What is needed is a theoretical framework that would enable us to define/compute the true generalisation and the accuracy of an estimation.
3. But how? Learn from others!

Theoretical Framework - Problem Formulation

1. In co-evolutionary learning, we consider the performance (quality) of a solution relative to other solutions.
2. This is achieved through interactions or comparisons between solutions.
3. This can be framed in the context of “game-playing”.

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- **Examples of Generalisation Framework**
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- Reading Lists

Theoretical Framework - $G_i(j)$

1. Let us have two strategies *i* and *j*.
2. A game is played between these two strategies.
3. Let $G_i(j)$ be the game outcome of strategy *i* playing against strategy *j*.
4. Similarly, let $G_j(i)$ be the outcome of strategy *j* playing against strategy *i*.
5. Strategy *i* is said to solve the test provided by strategy *j* if $G_i(j) \geq G_j(i)$.

   Strategy *i* wins against *j* if $G_i(j) > G_j(i)$.

True Generalisation Performance

1. Given a co-evolved strategy *i*, let test strategies *j* be obtained from strategy space S. The true generalisation performance of strategy *i*, G_i, is:

$$G_i = E_{P_1(j)}[G_i(j)] = \int_{\mathcal{S}} G_i(j) P_1(j) dj, \qquad (2)$$

   where G_i is the expectation of strategy *i*'s performance against *j*, $G_i(j)$, w.r.t. distribution $P_1(j)$ over strategy space S.

2. A simplified form:

$$G_i = \frac{1}{M} \sum_{j}^{M} G_i(j), \qquad (3)$$

   which is simply its average performance against all strategies *j*.

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- **Estimating Generalisation in Co-evolutionary Learning**
- Conclusions
- Reading Lists

Estimating Generalisation Performance

1. In reality, it is very hard or even impossible to compute the true generalisation performance using either Equation (2) or Equation (3).
2. We have to reply on estimation using a random sample of N test strategies *j*, S_N.
3. The estimated generalisation performance of strategy *i* is given by:

$$\hat{G}_i(S_N) = \frac{1}{N} \sum_{j \in S_N} G_i(j), \tag{4}$$

where S_N is the sample of N test strategies randomly drawn from S (we'll use notation

$$\hat{G}_i \ for \ \hat{G}_i(S_N)$$

How Good Is the Estimation

1. **We want to know how accurate the estimate $\hat{G}_i$ is compared to G_i, i.e., how small $|\hat{G}_i - G_i|$ is.**

2. **We don't know G_i, so we can't compute $|\hat{G}_i - G_i|$. So frustrating!**

3. **Fortunately, we can make a statistical claim as to how confident we are that $|\hat{G}_i - G_i| \leq \epsilon$.**

Chebyshev's Theorem

Theorem

Chebyshev's Theorem Consider a random variable U distributed according to the probability density $p(u)$. Given a positive number $a > 0$, we can bound the probability that $U \leq -a$ or $U > a$, i.e., the probability that U falls outside $[-a, +a]$, by

$$P(|U| \geq a) \leq \frac{E[U^2]}{a^2},$$

where $E[U^2]$ is the mean of the new random variable $V = U^2$ with respect to p.

Boris V Gnedenko. Theory of probability. Routledge, 2017

Bound for Estimation Accuracy

Applying Chebyshev's Theorem, we derive the following:

$$P(|\hat{G}_i - G_i| \geq \epsilon) \leq \frac{\sigma_i^2}{N\epsilon^2} \tag{5}$$

More Usable Bound

1. **In general, for the random variable $G_i(j)$ distributed over the interval $[G_{MIN}, G_{MAX}]$, $\sigma_{MAX} = (G_{MIN} + G_{MAX})/2$.**
2. **With this, we obtain the following lemma:**

Lemma

For a strategy i, let $\hat{G}_i$ be the estimated generalisation performance with respect to N random test strategies and G_i be the true generalisation performance. Consider the absolute difference $|\hat{G}_i - G_i|$, which is a random variable with distribution P_N taken on a compact interval $[G_{MIN}, G_{MAX}]$ of length $R = G_{MAX} - G_{MIN}$. Then, for any positive number $\epsilon > 0$:

$$P_N(|\hat{G}_i - G_i| \geq \epsilon) \leq \frac{R^2}{4N\epsilon^2}. \tag{6}$$

Harold I Jacobson. "The maximum variance of restricted unimodal distributions". In: The annals of mathematical statistics 40.5 (1969), pp. 1746–1752

Generality of the Generalisation Framework

The framework is extremely general.

1. The framework is independent of the complexity of the game, since it is independent of the size of the strategy space, and independent of the strategy distribution in the strategy space.
2. Both G_{MAX} and G_{MIN} are often known a priori as they are defined by the game, which means that we can always obtain an upper bound using (6).
3. The framework is independent of learning algorithms since the bound holds for any strategy in the strategy space.

Estimation Accuracy: An Illustration

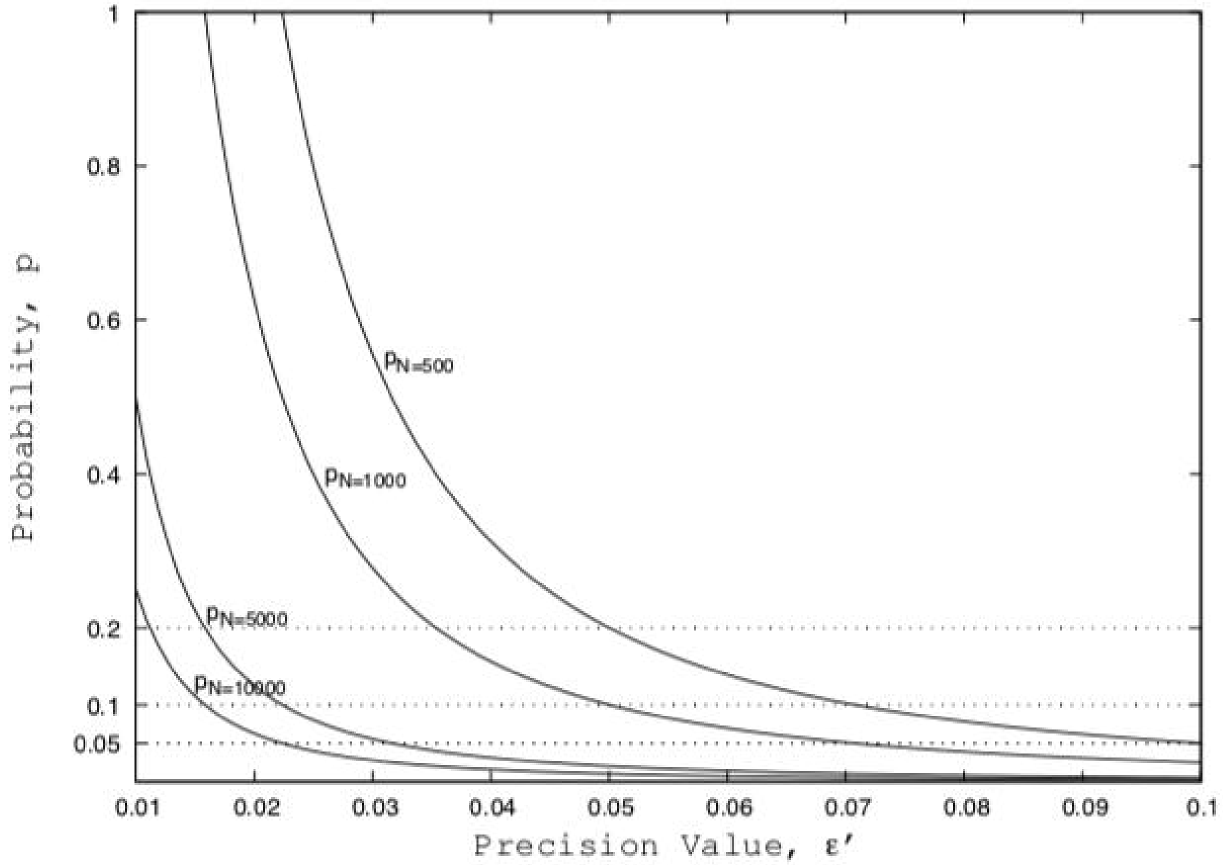


Figure 12: Equation (6) can be simplified to $P(|D_N|' \geq \epsilon') \leq \frac{1}{4N\epsilon'^2}$, where $\epsilon' = \epsilon/R$ and $|D_N|' = |\hat{G}_i - G_i|/R$. This figure shows the relationship between P_N for different N and precision ϵ' in $[0.01, 0.1]$.

Can We Obtain Tighter Bounds?

1. Chebyshev's bound is not very tight.
2. Furthermore, in Equation (5), we bound σ^2_i using R^2 to obtain Equation (6) (assuming the worst-case).
3. It is possible to find a tighter upper bound for σ^2_i .

Recall What We Did

Consider a random variable X **with the underlying distribution** P_X **over a real interval** $[a, b]$**, i.e.,** $X \in [a, b]$**. For** N **realisations** $x_1, x_2, \ldots, x_N$**, the empirical mean is** $\hat{E}_{P_X}[X] = \hat{\mu}_N = \frac{1}{N}\sum_{j=1}^{N} x_j$ **while the true mean is** $E_{P_X}[X]$**. The true variance is** $\sigma^2 = E_{P_X}[(X - E_{P_X}[X])^2]$. **Applying Chebyshev's Theorem gives us:**

$$P(|\hat{\mu}_N - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2} \leq \frac{R_X^2}{4N\epsilon^2},$$

where the maximum variance for a random variable over $[a, b]$ **is** $\sigma_{MAX}^2 = R_X^2/4$ **with** $R_X = b - a$**.**

A Tighter Bound on Estimation Accuracy

Lemma

For a strategy i, consider two independent non-overlapping sets of N test strategies: T_1 and T_2, where $T_1 \cap T_2 = \emptyset$ and $|T_1| = |T_2| = N$. The first set is used to estimate the generalisation performance $\hat{G}_i(T_1) = \frac{1}{N}\sum_{j \in T_1} G_i(j)$. The second set is used to estimate the variance $\hat{\sigma}^2_{N,U} = \frac{1}{N}\sum_{j \in T_2}(G_i(j) - \hat{G}_i(T_2) - \epsilon)^2$, for some positive number $\epsilon > 0$, where $\hat{G}_i(T_2) = \frac{1}{N}\sum_{j \in T_2} G_i(j)$. Then, for $\delta > 0$ with probability at least $c_1 c_2 = (1 - \frac{R^2}{4N\epsilon^2})(1 - \frac{R^4}{4N\delta^2})$, the following inequality holds:

$$P_N(|\hat{G}_i(T_1) - G_i| \geq \epsilon) \leq \frac{\hat{\sigma}^2_{N,U} + \delta}{N\epsilon^2}. \tag{7}$$

Measuring Generalisation Performance for IPD

1. **Let $g(i, j)$ be the average payoff per move to strategy i when it plays an IPD game with strategy j.**

2. **Generalisation performance in terms of the number of wins based on individual game outcomes:**

$$G_W(i,j) = \begin{cases} C_{WIN} & for\ g(i,j) > g(j,i), \\ C_{LOSE} & otherwise, \end{cases}$$

where $C_{WIN} > C_{LOSE}$.

3. **Generalisation performance in terms of average payoff:**

$$G_A(i,j) = g(i,j).$$

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- **Conclusions**
- Reading Lists

Conclusions

1. We have presented a theoretical framework for measuring generalisation performance rigorously in co-evolutionary learning. For the first time, quantitative analysis of generalisation performance of any co-evolutionary learning system can be performed.
2. The framework is extremely general and independent of any games, distributions and algorithms.
3. The theoretical framework can be applied to concrete games and algorithms, and estimate the generalisation performances.
4. Empirical results show that a small sample is usually good enough in estimating the generalisation performance.
5. More details: Siang Yew Chong, Peter Tino, and Xin Yao. “Measuring generalization performance in coevolutionary learning”. In: IEEE Transactions on Evolutionary Computation 12.4 (2008), pp. 479–505

Outline of This Lecture

- Introduction
- Strategy Games
- Co-evolutionary Learning of Game-playing Strategies
- Theoretical Framework of Generalisation in Co-evolutionary Learning
- Examples of Generalisation Framework
- Estimating Generalisation in Co-evolutionary Learning
- Conclusions
- **Reading Lists**

References for This Lecture I

1. Siang Yew Chong, Peter Tino, and Xin Yao. “Measuring generalization performance in coevolutionary learning”. In: IEEE Transactions on Evolutionary Computation 12.4 (2008), pp. 479–505.
2. Andrew M Colman. Game theory and experimental games: The study of strategic interaction. Elsevier, 2016.
3. Paul J Darwen and Xin Yao. “On evolving robust strategies for iterated prisoner’s dilemma”. In: Progress in Evolutionary Computation. Springer, 1993, pp. 276–292.
4. James H Davis, Patrick R Laughlin, and Samuel S Komorita. “The social psychology of small groups: Cooperative and mixed-motive interaction”. In: Annual review of Psychology 27.1 (1976), pp. 501–541.
5. Natalie S Glance and Bernardo A Huberman. “The dynamics of social dilemmas”. In: Scientific American 270.3 (1994), pp. 76–81.
6. Natalie S Glance and Bernardo A Huberman. “The outbreak of cooperation”. In:Journal of Mathematical sociology 17.4 (1993), pp. 281–302.

References for This Lecture II

7. Boris V Gnedenko. Theory of probability. Routledge, 2017.
8. Harold I Jacobson. “The maximum variance of restricted unimodal distributions”. In: The annals of mathematical statistics 40.5 (1969), pp. 1746–1752.
9. Xin Yao and Paul J Darwen. “An experimental study of N-person iterated prisoner’s dilemma games“. In: Progress in evolutionary computation. Springer, 1993, pp. 90–108.

Reading List for Next Lecture

1. H. Wang, L. Jiao, and X. Yao, “Two arch2: An improved two-archive algorithm for many-objective optimization,” IEEE Transactions on Evolutionary Computation, vol. 19, no. 4, pp. 524–541, 2015.
2. H. Chen and X. Yao, “Multiobjective neural network ensembles based on regularized negative correlation learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 12, pp. 1738–1751, 2010.