

ノイズ入り囚人のジレンマトーナメント大学対抗戦

矢野周一	富田勝	星野力	伊東秀昭
慶應義塾大学	慶應義塾大学	筑波大学	筑波大学
環境情報学部	Carnegie Mellon University		工学システム学類

概要

1994年10月1日に行なわれた「ノイズ入り囚人のジレンマトーナメント大学対抗戦」の結果を報告する。囚人のジレンマは非ゼロ和二人ゲームで、お互いに協調することによって両者とも利得を得ることができるが、一方が裏切ることによって相手に被害を与えつつも自分はより高い利得を得ることができる、というゲームである。歴史的にはアクセルロッド主催のトーナメントが有名であるが、本大会では「ノイズ入り」ルールで行なった。すなわち、プレイヤーが協調を意図しても、 ϵ %の確率で裏切りを実行してしまい、同様に裏切りを意図していても、 ϵ %の確率で協調してしまう。21戦略が集まり、予選、本戦とリーグを分けて優勝戦略を決定した。

Inter College Tournaments of Prisoner's Dilemma Game with Noise

Suichi Yano	Masaru Tomita	Tsutomu Hoshino	Hideaki Ito
Department of	Keio University	University of Tsukuba	Institute of
Environmental Information	Carnegie Mellon University		Engineering Mechanics
Keio University			University of Tsukuba

abstract

The following are the results of the Prisoner's Dilemma Game tournament held on October 1, 1994. In this Prisoner's Dilemma Game, each player has a choice of 'Cooperate' or 'Defect', and if both players cooperate, they both benefit, but if one player defects the other, the defector can get even higher payoff by damaging the other. Axelrod first organized a computer tournament of PD in 1980. We expand the model of the game with 'noise', i.e., each player's move may not always be made as intended; with the probability of ϵ %, Defect may become Cooperate, and Cooperate may become Defect. The 21 strategies entered were divided into 3 preliminary leagues, then a final league was made with the strategies which finished first or second.

1 はじめに

囚人のジレンマ問題はゲームの理論として数多く研究されている。

二人のプレイヤーがいて、それぞれ「協調」「裏切り」という2種類の行動を選ぶことができる。二人とも協調すると高い点が得られるし、二人とも裏切ると低い点しか得られない。二人の行動が異なった場合、裏切った方は最も高い点が得られる。しかし協調した方は点を得ることができない。このような状況を考えた時にどういう行動をとるのが最も有効かを考える問題である。

1回限りの囚人のジレンマゲームでは、裏切るのが賢明な戦略である。しかし繰り返しこれを行なう場合は、裏切り合って共倒れになるよりもお互いに協調して得点を稼いだ方が賢明である。

繰り返し囚人のジレンマゲームは、非ゼロ和二人ゲームで、特にアクセルロッドによる研究が有名である [1][2][3]。アクセルロッドは戦略を募集して2回トーナメントを行ったが、いずれも総当たり戦の結果、しっぺ返しと呼ばれる戦略が優勝した。この戦略は、最初の回は協調し、以後は前回の相手の手と同じ手を選ぶというものである。どちらの大会でも、上品な戦略(決して自分からは裏切らない戦略)が上位を占めた。上品な戦略同士が対戦すると、互いに協調しあい、利益を上げることが出来る。相手をだまして相手から搾取しようとする戦略は、結局自分と同じような搾取戦略と対戦すると得点を稼ぐことが出来ず、敗退した。トーナメントの中には、さらに点数をかせぐために工夫されたしっぺ返しの変形戦略もあったが、このような巧妙な改良はあまり役に立たなかった。アクセルロッドは、教訓として繰り返しジレンマでは妬みは禁物であるとしている。特に非ゼロサムゲームで大事なものは、相手よりたくさんの点を稼ごうとすることではなく、自分自身の利得を最大にしようとすることであると主張している。

また一般的に協調を C(Cooperation)、裏切りを D(Defection) というアルファベットであらわす。

2 トーナメントの方法

2.1 ルール

今回の囚人のジレンマゲームの得点行列を記す。

	協調	裏切り
協調	3,3	0,5
裏切り	5,0	1,1

2人のプレイヤーが協調しあうと各々3点ずつ得られる。お互いに裏切りあうと1点ずつ得られる。2人の行動が違う場合、協調した方はだまされて0点、裏切った方は最も高い5点を得ることができる。

2.2 ノイズについて

今回のトーナメントでは、ノイズを入れるというルールを拡張を行った。ノイズとは、ある確率でプレイヤーの意志に反した手を出す仕組みである。ノイズは通信路に入るので、自分の意図した手が間違っただけで伝わったり相手の手を間違っただけで認識したりする。自分の手はノイズが入って相手に伝わるし、相手の手もノイズが入ったものを受け取る。自分の出した手が相手にどう伝わったかは分からないし、逆に相手の手が相手の本心によるものかも知れない。

また、点数計算にはノイズが入った後の手を用いる。プレイヤーが正確な得点を知る事はできない。

2.3 トーナメントの運営

今回のトーナメントは、

参加者から戦略を募集（C言語） → 対戦プログラムによる点数計算・順位付け → 結果の告知

という流れで行った。

点数計算・順位付けは1994年10月1日に慶應義塾大学湘南藤沢キャンパスで、人工生命研究会の月例会として行なわれた。

2.4 システムについて

対戦システムは、C言語で作成されたクライアント・サーバシステムを用いた。これは独立したプロセス間の通信を基本として、プレイヤーがクライアントに、ゲームマスターがサーバとなって動く。対戦結果がXウィンドウでグラフと共に逐一表示される。このシステムで、参加者は戦略をC言語でコーディングして登録した。

この大会に参加した戦略は、インターネットで募集して集まった21種類の戦略である。詳細は付録に示す。

3 トーナメントの結果

トーナメントは予選リーグと決勝リーグに分けて行われた。まず21個の戦略を7つずつA、B、Cという3つの予選リーグに分けた。それぞれのリーグで総合得点の高い2戦略ずつが決勝リーグにすんだ。

対戦回数やノイズのパーセンテージは大会当日に決められた。今回は200ターン・ノイズ率5%となった。また、自分自身との対戦は行わなかった。

3.1 予選の結果

Aリーグ

	AKIYAMA1	IKEGAMI2	MIRINBOSHI	OZAKI	FUKUSHIMA	GOGO	GRET
AKIYAMA1		428	572	502	543	442	517
IKEGAMI2	383		384	325	440	402	380
MIRINBOSHI	602	339		571	463	445	613
OZAKI	517	480	441		696	510	431
FUKUSHIMA	593	195	528	336		530	592
GOGO	422	462	305	480	515		462
GRET	517	400	553	431	557	442	
総得点	3034	2304	2783	2645	3214	2771	2995
順位（総得点）	2	7	4	6	1	5	3

AリーグではFUKUSHIMA戦略が優勝した。決勝に進んだのはFUKUSHIMAとAKIYAMA1だった。

FUKUSHIMAはどの戦略からも点を取っており、大負けをしていない。AKIYAMA1、GOGOやGRETとは協調している。MIRINBOSHIには少し負けているものの、OZAKI戦略やGRET戦略から点を稼いでいる。さらにIKEGAMI2戦略に大差で勝っている。OZAKI戦略に対する得点力が、他の戦略と差をつけ優勝の原動力になったようだ。

AKIYAMA1も大負けをせずに点を稼いでいる。

Bリーグ

	IKEGAMI1	AKIYAMA3	NASHIDA	JAMIL	KAWAMURA	KOMOTO	OGAWA
IKEGAMI1		357	463	348	488	452	324
AKIYAMA3	407		273	521	575	570	585
NASHIDA	148	218		220	233	225	138
JAMIL	368	511	280		555	567	564
KAWAMURA	158	605	233	595		654	578
KOMOTO	187	600	250	592	374		592
OGAWA	599	600	569	609	593	587	
総得点	1867	2891	2068	2885	2818	3055	2781
順位(総得点)	7	2	6	3	4	1	5

Bリーグは KOMOTO、AKIYAMA3 という順位になった。AKIYAMA 戦略は 2 つとも決勝に進出したことになる。

KOMOTO は、NASHIDA 戦略と裏切りあって泥試合になり、点を伸ばせていないが、それ以外の戦略との対戦で順当に得点している。AKIYAMA3、JAMIL、OGAWA とは協調しており、KAWAMURA からは搾取している。この KAWAMURA に対する得点力が、優勝の鍵になったようだ。

AKIYAMA3 も同じように NASHIDA 戦略以外の戦略から点を稼いでいる。しかしこの戦略は大勝ちがなくて、まさに協調して点をのばしている。

Cリーグ

	TOQUEMON	SABUROTURBO	YOSHIKAWA	SAITO	SUGI	YOSHINAGA	CCD
TOQUEMON		459	438	443	360	537	423
SABUROTURBO	439		550	579	570	602	642
YOSHIKAWA	473	475		489	510	350	491
SAITO	368	579	544		374	411	401
SUGI	735	600	640	729		597	712
YOSHINAGA	212	572	235	231	577		160
CCD	543	442	536	581	401	665	
総得点	2770	3127	2943	3052	2792	3162	2829
順位(総得点)	7	2	4	3	6	1	5

Cリーグは YOSHINAGA と SABUROTURBO が決勝に進出した。

YOSHINAGA は、YOSHIKAWA 戦略と共倒れになっているようだが、TOQUEMON や SAITO、CCD から点を稼ぎ、その他の戦略とはうまく協調している。

SABUROTURBO は大負けも大勝ちもせず、コンスタントに点を稼いだようだ。

3.2 決勝戦の結果

各リーグを勝ち抜いた、合計 6 戦略で新たにリーグを組み、対戦させた。

決勝リーグ

	FUKUSHIMA	AKIYAMA1	KOMOTO	AKIYAMA3	YOSHINAGA	SABUROTURBO
FUKUSHIMA		513	281	550	270	530
AKIYAMA1	503		559	581	569	565
KOMOTO	241	614		589	104	589
AKIYAMA3	540	556	579		558	566
YOSHINAGA	260	609	894	603		582
SABUROTURBO	560	605	574	601	592	
総得点	2104	2897	2887	2924	2093	2832
順位(総得点)	5	2	3	1	6	4

決勝リーグはかなり混戦だった。FUKUSHIMA と KOMOTO、YOSHINAGA と FUKUSHIMA、YOSHINAGA と KOMOTO が潰しあってしまった。それで順当に協調した AKIYAMA1、AKIYAMA3 が高得点をマークした。

ということで、「AKIYAMA3」が優勝、「AKIYAMA1」が準優勝、3位は KOMOTO 戦略だった。KOMOTO 戦略は潰しあいもしたが、YOSHINAGA から大量得点を奪い、3位に食い込んだ。

AKIYAMA 戦略は、しっぺ返し (TFT) と堪忍袋 (TF2T) の混合戦略である。優勝した「AKIYAMA3」は 3/5 の確率で TFT、2/5 の確率で TF2T を出す。

4 考察

それぞれのリーグで上位に食い込んだ戦略は、それぞれ大負けをしていない。これはアクセルロッドの、自分の利益を最大にするように振舞えという主張を裏付けている。相手を出し抜くよりも協調してお互いに利益を得るということだ。

最終的に勝った戦略は TFT と TF2T の混合戦略だった。ノイズ入りの環境では、戦略にいかに寛容性を持たせるかがポイントになってくる。アクセルロッドの大会で優勝した TFT はノイズに弱い。一度ノイズが入ると、確実に反応して裏切り返してしまう。もしも相手が TFT や、TFT によく似た戦略だったら、お互いに潰しあってしまう。従ってすぐに裏切り返さないよう、戦略に寛容性を持たせることが高得点への近道である。

後日、200回・ノイズ5%、自分自身との対戦はやらない総当たり戦を10回行なって総得点を出した。その結果、大幅に順位が入れ替わった。ベスト5は次のような戦略になった。(全結果は付録に示す。)

10860:Fukushima 9942:Akiyama1 9726:Ikegami2 9541:Akiyama3 9519:Ikegami1

トーナメントでは決勝リーグで5位に終わった FUKUSHIMA が、総当たり戦で優勝した。これは、学習する FUKUSHIMA 戦略の強さを裏付けている。慶應大学だけで行なったトーナメント [4] でもそうだったのだが、今回も上品でない戦略が多かったので、強さを発揮したのではないだろうか。

また、予選で低迷した IKEGAMI 戦略が上位に進出した。これは戦略を GA で進化させたものである。

5 今後の課題

今回は自分自身との戦いを行わなかった。従って自分自身同士で点を稼ぐことがなかったといえる。これは裏切りがちな戦略に有利にはたらいたのではないか。なぜならば全て裏切る戦略などは自分自身同士では、低い得点しか上げられないからである。協調行動の進化を見るためには、やはり同じ戦略同士が協調できる環境を考えてみたいと思う。

リーグを分割して行ったので対戦相手による運・不運があったと思われる。実際に IKEGAMI 戦略は分割リーグではふるわなかったが、総当たり戦では上位に進出した。この点も改善の余地があると思われる。

6 謝辞

筑波大学星野研究室の方々、慶應大学富田研究会の方々、および関係者に感謝いたします。

参考文献

- [1] Axelrod, The Evolution of Cooperation, Basic Books, 1984.
- [2] W.D.Hamilton: The Evolution of Cooperation, Science, vol.211, 27 March 1981, pp.1390-1396.
- [3] R.Axelrod, D.Dion, The Further Evolution of Cooperation, Science, vol. 242, 9 Dec 1988, pp.1385-1390.
- [4] 矢野周一, 城戸隆, 富田勝, ノイズ入り四人のジレンマトーナメント, Game Programming Workshop, Nov 1994.

7 付録 (参加戦略の説明・製作者)

製作者名、プログラム名、アルゴリズムの説明です。

池上高志 (いけがみたかし)、IKEGAMI1 & 2

戦略は最近の過去2回の対戦で、1回前にお互い協力していれば次は協力。お互い裏切っていた場合さらに1回前にお互い裏切っているかお互い協力していれば協調する。このパターン以外は裏切るというものです。最初の1、2手は裏切りです。

IKEGAMI1 と IKEGAMI2 の違いは、IKEGAMI2 の場合は IKEGAMI1 より協力するパターンがひとつ少なく、1回前にお互い協力していれば IKEGAMI1 なら協力ですが、IKEGAMI2 ではさらに1回前をみて、自分が裏切り相手が協力の場合は裏切ります。多分これは意味がないので、基本的に同じ戦略です。

千田 芳裕 (ちだ よしひろ)、SABURROTURBO

基本的には過去3回の相手の手を見て多い方の手を返すだけだが、連続して裏切らない機構と、相手が無反応型(ランダム・周期戦略など)だと感知すると裏切りに転じる機構を持っている。

連続して裏切らない機構:

相互不信の状態に陥ったときこちらから相手を許すシステム。

自分の裏切りの回数をスタックに記憶しておく(裏切ると増え、協調すると減る)。裏切り回数があるレベル(ボーダーライン)に達すると、互いに裏切りの応酬になったとみなし、過去の相手の裏切りを忘れてあげる。

しかもボーダーライン自体も、このシステムが頻繁に作動するようだと上昇し(なかなか許さなくなる)、あまり作動しないと下降する(すぐに許すようになる)。

無反応を感知して裏切る機構:

こちらが出した手に対して相手が反応しないようだと裏切るシステム。

自分が裏切りの手を出したとき、相手がどう反応したかを過去20回分キューに記憶しておく。9回以上が協調ならば無反応だと判断して裏切る。

これは「連続して裏切らない機構」より優先されるので裏切り続けることもありえる。しかし相手がこちらの裏切りに反応して裏切り返せばこのシステムは停止するので、裏切りの応酬にはならないはずである。

千田 芳裕 (ちだ よしひろ)、MIRINBOSHI

基本は Tft だが相手の出方によって「信用度」を増減させ、「信用度」に従って自分の戦略を変える。

相手の出方は「Tft率」で判定する。これは、それまでに自分が出した手に対し相手も同じ手を返した率で、ターンごとに計算する。信用度は初め100%で、「Tft率」に近づくように徐々に調整する。

自分の行動は次のように決める:

(1) 信用度が70%以上で、しかも現在までのTft率より低い時協調

(2) 信用度が30%以上の時 Tft

(2) 信用度が30%未満の時裏切り

信用度は自分が裏切ると減少し、自分が協調した時相手も協調すると増加する。

吉川恵健 (よしかわえけん)、YOSHIKAWA (GTFT+JOSS) #前者はコンテスト時の名前、でも本当は括弧内の名前を付けていました。

基本的には Tft 戦略だが、15%の確率で前回相手の出した手と逆の手を出す。つまり普段は Tft であるが、ときどき相手の裏切りを許容したり (GenerousTft)、相手の協調を裏切ったり (JOSS) するという戦略である。GTFT はノイズ下で泥試合から脱出するために有利であり、JOSS は時々食い逃げできそうなことから、両者の利点を併せる目的で作成した。

福島 智和(ふくしま ともかず)、FUKUSHIMA

まず、自分がおかれた状況(前の回の様子)を
相手が協調自分も協調・相手が裏切り自分は協調・相手が協調自分は裏切り・相手が裏切り自分も裏切り。
の4種類に分ける。そして全ての履歴を参照し、現在の状況と同じ状況の回を全てさがす。そして、そのつぎ
の回の自分の手ごとに以降三回の自分の得点の平均をとり、得点の高い方の手を出す。得点と同じ時は裏切
り。データが無い場合はデータが無い方の手を出してデータ収集をする。初めの10回はしっぺがえし。

北川 宏和(きたがわ ひろかず)、GOGO

基本的に相手戦略に周期性があるか検索して、周期戦略と断定できたら以後は裏切り続ける。

詳しくは20回から5回ごとに相手に周期性があるかどうかを探す。相手の過去の戦略を2~5ずつに切
て10個並べる。10に満たない場合は検索は行わない。それで、ノイズを考えてある一定個数以上同じリス
トが現れていたならばそれは周期であるとする。それが10回以上連続して現れた場合、完全に周期戦略であると
断定する。逆に5回以上連続で周期性が認められなかった場合、非周期戦略であるとみなし、以後はRET1
と同じ振舞いをする。

城戸 隆(きど たかし)、GRET

前回相手が協調したときには100%で協調、裏切ったときには10%で協調。

飯田 一紀(いいだ かずき)、JAMIL

ちょっと反応の遅いしっぺ返し。詳しくは相手の3回前の手を出す、というもの。

川村 親平(かわむら しんべい)、KAWAMURA

一回めはcを出し、あとは相手の手を見てCが多ければC、Dが多ければD回数だったらCを出す。

河本 信昭(こうもと のぶあき)、KOMOTO

- 1、20回までシッペガエシ。
- 2、まだ1度もD(C)を出していないような相手にはD(C)を出す。確率の比較が出来るようにするため。
- 3、あとは(自分がCを出した場合に相手がCを出す確率)と、(自分がDを出した場合に相手がCを出す確率)
を比較し、前者の方が「より」高い場合はCを それ以外はDを(勿論おなじ場合はD)出す。

小川 智明(おがわ ともあき)、OGAWA

自分が気持ちいいと協調し、気分が悪いと裏切る。

自分裏切り、相手協調 = うれしい = 協調度10点

自分協調、相手協調 = まあいい = 協調度5点

自分裏切り、相手裏切り = 少し嫌 = 協調度3点

自分協調、相手裏切り = 気分が悪い = 協調度-1点

10回前までの履歴を調べ、点数の合計で、次の手を決める。最初の10回はしっぺ返し。

斎藤 輪太郎(さいとう りんたろう)、SAITO

最初の10回はCとDをてきとうに出して様子を見る。

次に自分の履歴と相手の履歴を見て、80%以上の割合で自分の出した手を相手が出しているなら相手はしっ
ぺ返しの戦略を取っていると判断しCを出す。

相手がしっぺ返しの戦略をとっていると判断しなかった場合はDを4回、Cを1回出し、その後しっぺ返
しの戦略を5回取る。

以降相手がしっぺ返しの戦略を取っていると判断したときはC、そうでなければ、Dを4回Cを1回としっ
ぺ返し5回のうち得点の多い方の戦略を取る。

杉山 雄一(すぎやま ゆういち)、SUGI

この戦略ははじめ2回連続・もしくは一定の割合以上裏切られない限り報復しないが、そうなるとしてっ返し戦略に切り替えます。それでもしつこく裏切ってくる相手には全面背信行為をとりますが、相手に反省の色がみられればしっぺ返しに戻る。

進藤 隆志(しんどう たかし)、CCD

ccdccdccc... と、ccd を繰り返す周期戦略。

吉永 潤(よしなが じゅん)、YOSHINAGA

20 回中 5 回以上 D をだしたら ALLD。そうでない場合は過去相手が出したのと同確率で C または D を出すつもりだっ
た。しかし D のカウンターを最初の 20 回のなかにいれてしまっていたので、相手が 21 回目以降にだした D をカウントできずに計算してしまい、たとえば 150 回目なら 0~4/150 くらいでしか、D をださなくなってしまった。

秋山 英三、AKIYAMA1、AKIYAMA3

「AKIYAMA1」は 1 1 回のうち 2 回周期的に Tf2T が入るようになってます。「AKIYAMA3」はランダムに 2/5(? だったと思います) の確率で Tf2T が入るようにしました。

徳永幸彦、toquemon

まず、自分と相手の出した手は全て配列 my_history[] と your_history[] に記憶されます。この配列の他に strategy[] という配列があります。最初この配列は長さ 1 に設定されており、1 回前の記憶しか戦略として参照できません。つまりこの配列の長さが、記憶の長さに対応するわけです。戦略の記憶が 1 の場合、もし 1 が立っていたら、自分と相手が 1 回前に出した手の AND を戦略にします。例えば自分が前の turn で 1 (DEFECT) 相手が 0 (COOPERATE) だったら、0 を返し、自分も相手も 0 または 1 だったら 1 を返します。もし戦略に 0 が立っていた場合には、自分と相手の手の XOR を返します。ですから、自分が 1 で相手が 0 だったら 1 を返しますし、自分も相手も 0 または 1 だったら 0 を返します。

1 回前の turn で自分が 0 相手が 1 の時(つまり、自分が搾取された時)には、戦略の変更をします。まず戦略の長さがランダムに伸び縮みます。但し増減幅は 1 で、伸びる方向にバイアスがかかっています。戦略の長さが決定されると、それによって戦略の配列に 0 か 1 がランダムに割り当てられます。

記憶が深くなっていった場合は、最終的な手の決定は次のように計算します。今、下のような対戦で過去 5 回戦い、戦略が 110 になっていたとします。この戦略を過去の方向に延ばして、自分と相手の手から AND または XOR を計算します。計算結果の和(この場合 0+1+0=1)を 2 で割った余りを、次に出す手とします。

turn	1	2	3	4	5	turn	1	2	3	4	5
my history	0	0	1	0	1	strategy			0	1	1
your history	0	1	1	0	0	ANDorXOR			0	1	0

(0+1+0)/2=0...1----> my_next_action

無田 廣之 (なした ひろゆき)、パブロフ改

パブロフ戦略(NATURE, VOL-364, 1-JULY-1993, pp.56-58)の改悪版。相手に勝つことだけを目的としたアルゴリズム

尾崎 ゆういち、TF2T 改

裏切る条件は次の 2 つの条件のうち、どちらか(または両方)が成立した時

1: 相手が 2 度連続裏切ったとき 2: 相手が 4 回連続協調したとき それ以外は協調。

総あたり戦の結果

10860:Fukushima	9942:Akiyama1	9726:Ikegami2	9541:Akiyama3	9519:Ikegami1	9441:Sugi	9362:Yoshinaga
9335:Mirinboshi	9292:Yoshikawa	9248:Gret	9244:Saburoturbo	9236:Saito	9188:Ozaki	9167:Nashida2
8928:Jamil	8755:Ccd	8416:Toquemon	8124:Ogawa	7608:Koumoto	6892:Gogo	6744:Kawamura