

Hedging Algorithms and Repeated Matrix Games

Bruno Bouzy, Marc Métivier and Damien Pellier

LIPADE
Université Paris Descartes



Workshop on Machine Learning and Data Mining in Games
Athens, Greece
September 9th 2011

Outline

- 1 Hedging algorithms
- 2 Multi-agent learning (MAL) and Repeated Matrix Games (RMG)
- 3 Basic players and settings
- 4 Key idea, example
- 5 Experimental proof
- 6 Several levels
- 7 Conclusion and future work

Hedging algorithms

- An hedging algorithm is made up with
 - a top-level algorithm
 - a set of basic algorithms (or experts)
- To make its decision the hedging algorithm
 - asks the top-level algorithm to choose a basic algorithm
 - the chosen algorithm makes the decision
- Related work
 - Weighted majority algorithm (Littlestone and Warmuth 1994)
 - Aggregating algorithm (Vovk 1998)
 - Multiplicative weight (Freund and Schapire 1999)
 - Hierarchical hedging algorithm (Chang and Kaelbling 2005)
 - Exploration-Exploitation Experts method (Pucci and Megiddo 2006)

Multi-agent learning (MAL)

- Multi-agent learning (MAL)
 - Several agents are learning simultaneously
 - The environment is modified by the other agents
- The 5 MAL agendas (Shoham and al. 2007)
 - Cognitive (describe human learning), Normative (use game theory), Computational (compute Nash equilibria)
 - Prescriptive (maximize cumulative return)
 - cooperative: the agents communicate between each other
 - non cooperative: the agents do not communicate between each other
- The prescriptive non cooperative agenda: mainly with RMG
 - Minimax QL, Friend and Foe QL, Bully (Littman 1994, 2001)
 - Awesome (Conitzer and Sandholm 2003)
 - M3, S (Crandall and Goodrich 2003, 2005)
 - Meta, Manipulator (Powers and Shoham 2004, 2005)

Repeated Matrix Games (RMG)

	c	d
a	2, 2	0, 0
b	0, 0	1, 1

coordination

	c	d
a	-3, 3	0, 0
b	-1, 1	-2, 2

competition

	coop	default
coop	3, 3	1, 4
default	4, 1	2, 2

prisoner dilemma

	foot	theater
foot	2, 1	0, 0
theater	0, 0	1, 2

battle of sexes

	c	d
a	1, 0	3, 2
b	2, 1	4, 0

stackelberg

	c	d
a	6, 6	2, 7
b	7, 2	0, 0

chicken

Related work in RMG

- Airiau and al 2007
 - 9 algorithms, 5 learners (only)
 - Best algorithm: Fictitious Play (FP)
 - No QL based algorithm, no bandit based algorithm
- Zawadzki 2005
 - 11 algorithms, 9 learners (AWESOME, QL, GIGA-WoLF, Meta ...)
 - QL is the best.
 - Nash-based algorithms inferior to a best response algorithm (QL).
 - No bandit based algorithm.
- Bouzy and Metivier 2010
 - 12 algorithms, 9 learners (M3, S, UCB, QL, FP, Exp3...)
 - 2 features do improve the learners
 - fixed-width window heuristic (w)
 - the state heuristic: state = previous joint action (s)
 - M3, UCB+s+w, QL+s+w are the best algorithms

Basic players

- Learning players
 - FP: Fictitious Play (Brown, 1951)
 - QL: Q-learning (Watkins & Dayan, 1992)
 - S: (Stimpson & Goodrich, 2003)
 - M3: M-Qubed (Crandall & Goodrich, 2005)
 - JR: (J. Robinson 1951)
 - HMC: (Hart & Mas-Colell 2000)
 - UCB: (Auer et al., 2002)
 - Exp3: (Auer et al., 2002b)
- Non learning players
 - Bully (Stone & Littman, 2001)
 - Minimax (J. Robinson 1951)
 - Greedy
 - Random

Settings

- An all-against-all tournament (AGAT) using a given RMG
 - Each player matches each player (once playing column, once playing row)
 - player evaluation = average return over the matches
- An experiment is a set of N AGAT
 - For each AGAT, a RMG is drawn at random (returns in $[-9, 9]$)
 - player evaluation = average return over the tournaments
 - the output of an experiment is a ranking

Precision over average values computed

- Randomness has 2 sources.
- Randomness at each repetition in the decision process of the players:
 - Returns in $[-9, +9]$.
 - About 100,000 repetitions per RMG match.
 - Standard deviation $\approx 0.003 \times 9 \approx 0.03$.
- Randomness to draw the RMG used in an AGAT:
 - Using the elimination principle: after each AGAT, the last player is eliminated (Bouzy and Metivier 2010).
 - Random variable X = average value of the last player of an AGAT.
 - Observed standard deviation of $X \approx 3$.
 - About 1,000 RMG per experiment.
 - Observed standard deviation of $\bar{X} \approx 0.03 \times 1.5 \approx 0.05$

Key idea, example (1/3), basic result

- 3 basic players: UCB (U), Greedy (G) and S compared on 100 RMG.
- The result is a **ranking: 1: U 28.7 2: S 27.0 3: G 24.5**
- The result can be detailed by a table.
- Each cell corresponds to a match between two players and contains the average returns obtained by the two players.

	U	G	S	rt	T
U	4.45 4.42	4.11 4.48	5.77 4.33	14.33	28.67
G	4.53 4.07	3.12 3.04	4.59 4.35	12.24	24.53
S	4.15 5.86	4.35 4.77	4.85 4.99	13.34	27.01
ct	14.35	12.29	13.67		

- The table corresponds to a matrix game called the **algorithm game**.
- The player must choose an algorithm which plays the experiment.

Key idea, example (2/3), expected virtual result

- Let H be a perfect virtual player at the algorithm game: its opponent being known, it chooses the best algorithm adequately.
- The expected result of a virtual experiment between U , G , S and H .

	U	G	S	H	rt	T
U	4.45 4.42	4.11 4.48	5.77 4.33	4.11 4.48	18.44	36.86
G	4.53 4.07	3.12 3.04	4.59 4.35	4.59 4.35	16.83	33.89
S	4.15 5.86	4.35 4.77	4.85 4.99	4.15 5.86	17.49	35.49
H	4.53 4.07	4.35 4.77	5.77 4.33	5.77 4.33	20.42	39.44
ct	18.42	17.06	18.00	19.02		

- Expected ranking: **1. H 39.4 2. U 36.9 3. S 35.5 4. G 33.9**
- H might be the best player.

Key idea, example (3/3), actual result

- Let H be Hedge(top=S, U, G, S)
- The actual result of an experiment between U, G, S and H.

	<i>U</i>	<i>G</i>	<i>S</i>	<i>H</i>	rt	T
<i>U</i>	4.42 4.39	4.11 4.47	5.89 4.26	5.30 4.75	19.73	39.30
<i>G</i>	4.51 4.07	3.18 3.07	4.70 4.38	1.12 5.77	13.51	27.21
<i>S</i>	4.12 5.84	4.36 4.68	4.78 4.93	4.36 5.37	17.62	35.57
<i>H</i>	4.68 5.26	5.61 1.49	5.29 4.38	4.92 4.93	20.49	41.32
ct	19.57	13.71	17.95	20.83		

- Actual ranking: **1. H 41.3 2. U 39.3 3. S 35.6 4. G 27.2**
- H is the actual best player.
- Actually the hedging principle works!**
- Explanation: H adapts quickly to choose the adequate player.

Experimental proof (1/3), quickly remove bad players

- Let $H = \text{Hedge}(\text{top}=A, B, C)$ be a hedging algorithm with A, B, C in the set of basic players.
- 12^3 combinations.
- Using the elimination principle (Bouzy and Metivier 2010).
- Observations:
 - **top=S** is mandatory.
 - Many combinations do not work at all.
 - Few combinations does not help much: $H(S, J, S)$, $H(S, S, S)$, $H(S, Q, S)$, $H(S, Q, Q)$.
 - Very few combinations helps: $H(S, M3, S)$, $H(S, M3, M3)$.
 - The best 2 significant combinations are **$H(S, M3, J)$** , **$H(S, U, M3)$** .

Experimental proof (2/3)

- $HSUM = H(S, U, M3)$ surpasses the basic players.

Table: Experiment with $HSUM = H(S, U, M3)$ in the league.

		Av. return
1	<i>HSUM</i>	4.88
2	<i>M3</i>	4.74
3	<i>U</i>	4.60
4	<i>J</i>	4.15
5	<i>S</i>	4.17
6	<i>B</i>	4.47
7	<i>Q</i>	4.49
8	<i>MinMax</i>	4.64
9	<i>R</i>	-2.78

Experimental proof (3/3)

- $HSUM = H(S, U, M3)$ surpasses the basic players enhanced with the w and s heuristics.

Table: Experiment with $Uw + s$, $M3$, $Jw + s$, S , $Qw + s$, and $H = Hedge(top = S, U, M3)$.

	Player	Av. return
1	$HSUM$	5.03
2	$Uw + s$	4.67
3	$M3$	4.57
4	$Jw + s$	4.47
5	S	3.80
6	$Qw + s$	3.58

- Better to use the hedging principle rather than adding the specific enhancements s and w .

Two levels (1/3)

- What is the effect of repeating the hedging principle twice?
- Let $HH = Hedge(S, U, G, S, Hedge(S, U, G, S))$

Table: The results of an experiment between U , G , S , H , and HH .

	U	G	S	H	HH	T
U	4.5 4.4	4.1 4.5	5.8 4.3	5.3 4.8	5.3 4.8	50.0
G	4.5 4.1	3.1 3.1	4.6 4.4	1.3 5.7	1.0 6.0	29.5
S	4.1 5.9	4.4 4.6	4.8 5.0	4.4 5.4	4.2 5.5	44.4
H	4.7 5.3	5.6 1.5	5.2 4.3	4.9 5.0	4.7 5.0	51.0
HH	4.6 5.3	5.8 1.3	5.3 4.4	4.9 4.9	4.8 4.9	51.7

- 2 levels are promising.

Two levels (2/3)

- Let $HHUMM = Hedge(top = S, HSUM, M3)$.

Table: Experiment with two levels in the league.

	Player	Av. ret.
1	<i>HHUMM</i>	5.04
2	<i>HSUM</i>	4.95
3	<i>M3</i>	4.46
4	<i>U</i>	4.44
5	<i>S</i>	3.79

- HHUMM surpasses HSUM and the basic players.

Two levels (3/3)

- HHUMM in the league of enhanced basic players.

Table: Experiment with *HSUM*, $Uw + s$, *M3*, $Jw + s$, *S*, $Qw + s$, and *HHUMM*.

	Player	Av. ret.
1	<i>HHUMM</i>	4.94
2	<i>HSUM</i>	4.88
3	$Uw + s$	4.72
4	<i>M3</i>	4.56
5	$Jw + s$	3.46
6	<i>S</i>	3.20
7	$Qw + s$	3.21

- HHUMM surpasses HSUM and the enhanced basic players.
- **Two levels work.**

Three levels

- Let $HHHUMM = Hedge(top = S, HHUMM, M3)$.

Table: Experiment with three levels in the league.

	Player	Av. ret.
1	<i>HHUMM</i>	5.08
2	<i>HHHUMM</i>	5.05
3	<i>HSUM</i>	5.13
4	<i>M3</i>	4.77
5	<i>U</i>	4.58
6	<i>S</i>	4.07

- HHHUMM surpasses HSUM and the basic players.
- However, HHHUMM does not surpass HHUMM.
- Three levels do not work.

Conclusion

- Summary
 - Many MAL algorithms playing RMG exists.
 - Playing RMG is a hard task to improve (Bouzy and Metivier 2010).
 - The basic players being given, hedging algorithms lead to a good final solution.
 - 2 instances of hedging algorithms are significantly better than all the basic algorithms.
 - It works indeed.
- Future work
 - Why $\text{top}=\text{S}$ is the single solution for the top algorithm ?
 - Try hedging algorithms with more than 2 or 3 basic algorithms.
 - Find out all the good solutions with a systematic approach (genetic algorithm?).
 - Use of hedging algorithms in other domains, or frameworks (stochastic games?).

Thank you for your attention!