Empirical Game-Theoretic Analysis for Practical Strategic Reasoning

Michael P. Wellman
University of Michigan

Workshop on Reasoning in Adversarial and Noncooperative Environments

18 Nov 2010

Real-World Games

- rich strategy space
  - strategy: obs* x time \(\rightarrow\) action
- severely incomplete information
  - interdependent types (signals)
  - info partially revealed over time
- analytic game-theoretic solutions few and far between

two approaches
1. analyze (stylized) approximations
   - one-shot, complete info...
2. empirical methods
   - simulation, statistics, machine learning, ...

complex dynamics and uncertainty
Empirical Game-Theoretic Analysis (EGTA)

- Game described *procedurally*, no directly usable analytical form
- Parametrize strategy space based on *agent architecture*
- Selectively explore strategy/profile space
- Induce game model (payoff function) from simulation data
TAC Supply Chain Mgmt Game

suppliers
- Pintel
- IMD
- Basus
- Macrostar
- Mec
- Queenmax
- Watergate
- Mintor

manufacturers
- Manufacturer 1
- Manufacturer 2
- Manufacturer 3
- Manufacturer 4
- Manufacturer 5
- Manufacturer 6

component RFQs
supplier offers
component orders

10 component types
16 PC types
220 simulation days
15 seconds per day

Two-Strategy Game (Unpreempted)
Two-Strategy Game (Unpreempted)

Three-Strategy Game: Deviations
TAC/SCM-06 Deviation Graph

CDA Deviation Graph

4 strategies:
GD, GDX, ZI, Kap
Ranking Strategies: TAC/SCM-07

<table>
<thead>
<tr>
<th>SCM-07 Tournament</th>
<th>SCM-07 EGTA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agent</strong></td>
<td><strong>Finals</strong></td>
</tr>
<tr>
<td>PhantAgent</td>
<td>8.67</td>
</tr>
<tr>
<td>TacTex</td>
<td>6.31</td>
</tr>
<tr>
<td>DeepMaize</td>
<td>5.45</td>
</tr>
<tr>
<td>Maxon</td>
<td>1.79</td>
</tr>
<tr>
<td>Tinhorn</td>
<td>1.34</td>
</tr>
<tr>
<td>CMieux</td>
<td>1.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Agent</strong></th>
<th><strong>NE Regret</strong></th>
<th><strong>Max Regret</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DM07 S [C07-9]</td>
<td>0.32</td>
<td>3.40</td>
</tr>
<tr>
<td>DM07 F [C07-34]</td>
<td>0</td>
<td>2.63</td>
</tr>
<tr>
<td>PH07</td>
<td>0</td>
<td>48.84</td>
</tr>
<tr>
<td>TT07 S</td>
<td>2.90</td>
<td>16.95</td>
</tr>
<tr>
<td>TT07 F</td>
<td>0</td>
<td>10.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Agent</strong></th>
<th><strong>NE Regret</strong></th>
<th><strong>Max Regret</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DM06 S</td>
<td>3.21</td>
<td>8.17</td>
</tr>
<tr>
<td>PH06</td>
<td>1.31</td>
<td>11.00</td>
</tr>
<tr>
<td>TT06</td>
<td>1.03</td>
<td>14.78</td>
</tr>
<tr>
<td>MR05</td>
<td>2.98</td>
<td>14.67</td>
</tr>
</tbody>
</table>

*from PR Jordan PhD Thesis, 2009*

Strategy Ranking (TAC Travel)

**Strategies ranked with respect to the final equilibrium context**

*from LJ Schwartzman PhD Thesis, 2009*
### DeepMaize-08 Design Exploration

<table>
<thead>
<tr>
<th>ID</th>
<th>Customer Dataset</th>
<th>Component Horizon Treatment</th>
<th>Controller</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCM05</td>
<td>SCM[06-07]</td>
<td>AIO</td>
</tr>
<tr>
<td>0</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3-5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>17-27</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>28-29</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 5.15: DeepMaize 08 tested feature matrix.

*from PR Jordan PhD Thesis, 2009*

### Iterative EGTA Process

![Diagram of Iterative EGTA Process]

- **JW, AAMAS-08**
  - Sampling Control Problem
  - Profile
  - Profile Space
  - Strategy Set
  - Add Strategy
  - Strategy Space
  - Strategy Exploration Problem
  - Game Analysis (NE)
  - Refine?
  - More Strategies
  - More Samples
  - End

- **JW, AAMAS-09**
  - Game Model Induction Problem
  - Empirical Game
  - Game Analysis

- **JSW, AAMAS-10**
  - Add Strategy
  - Strategy Space
  - Strategy Exploration Problem
Sampling Control Problem

• Revealed payoff model
  – sample provides exact payoff
  – minimum-regret-first search (MRFS)
    • attempts to refute best current candidate

• Noisy payoff model
  – sample drawn from payoff distribution
  – information gain search (IGS)
    • sample profile maximizing entropy difference wrt probability of being min-regret profile

Min-Regret-First Search

start (arbitrary)
Min-Regret Search

Select random deviation from current best profile
Min-Regret Search

evaluated
best

Min-Regret Search

evaluated
best
### Min-Regret Search

<table>
<thead>
<tr>
<th>(r1,c1)</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r1,c2)</td>
<td>5</td>
</tr>
<tr>
<td>(r2,c1)</td>
<td>3</td>
</tr>
<tr>
<td>(r3,c1)</td>
<td>7</td>
</tr>
<tr>
<td>(r1,c4)</td>
<td>1</td>
</tr>
<tr>
<td>(r2,c4)</td>
<td>1</td>
</tr>
</tbody>
</table>

![Evaluated and Best Positions](image)

**Evaluated** positions are marked in blue and **best** positions in green.
Min-Regret Search

<table>
<thead>
<tr>
<th>r1,c1</th>
<th>r1,c2</th>
<th>r2,c1</th>
<th>r3,c1</th>
<th>r1,c4</th>
<th>r2,c4</th>
<th>r2,c2</th>
<th>r2,c3</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,5</td>
<td>3,3</td>
<td>4,8</td>
<td>3,0</td>
<td>5,3</td>
<td>2,2</td>
<td>2,0</td>
<td>2,2</td>
</tr>
</tbody>
</table>

Profile $\varepsilon$-bound:

- $(r1,c1) = 3$
- $(r1,c2) = 5$
- $(r2,c1) = 4$
- $(r3,c1) = 7$
- $(r1,c4) = 1$
- $(r2,c4) = 5$
- $(r2,c2) = 0$
- $(r2,c3) = 8$
- $(r3,c2) = 6$
Min-Regret Search

|  | (r1,c1) | 3  |
|  | (r1,c2) | 5  |
|  | (r2,c1) | 4  |
|  | (r3,c1) | 7  |
|  | (r1,c4) | 1  |
|  | (r2,c4) | 5  |
|  | (r2,c2) | 0  |
|  | (r2,c3) | 8  |
|  | (r3,c2) | 6  |
|  | (r4,c2) | 6  |

Finding Approximate PSNE
Iterative EGTA Process

Construct Empirical Game

- Simplest approach: direct estimation
  - employ control variates and other variance reduction techniques

\[
\begin{align*}
(s_1, u(s_1)) \\
\vdots \\
(s_i, u(s_i))
\end{align*}
\]

Payoff data from selected profiles
Payoff Function Regression

\[ S_i = [0,1] \]

\[ \Downarrow \text{generate data (simulations)} \]

\[
\begin{array}{ccc}
0 & 0.5 & 1 \\
3,3 & 1,4 & 1,1 \\
4,1 & 2,2 & 4,1 \\
1,1 & 1,0 & 3,3 \\
\end{array}
\]

\[ \Downarrow \text{learn regression} \]

\[ \Downarrow \text{solve learned game} \]

\[ \text{eq} = (0.32,0.32) \]

Vorobeychik et al., ML 2007

Generalization Risk Approach

- Model variations
  - functional forms, relationship structures, parameters
  - strategy granularity

- Approach:
  - Treat candidate game model as a predictor for payoff data
  - Adopt loss function for predictor
  - Select model candidate minimizing expected loss

Cross Validation

Observation Data

Fold 1
Fold 2
Fold 3
Training
Validation
Sensitivity Analysis

Iterative EGTA Process

Sampling Control Problem
Profile Space
Profile
Select

Profile Space
Strategy Set
Add Strategy

Strategy Space
Strategy Exploration

Game Model Induction Problem
Payoff Data
Empirical Game

Game Analysis (NE)
Refine?

More Strategies
More Samples

JSW, AAMAS-10

End

M. Wellman
Nov-20-10
Learning New Strategies: EGTA+RL

CDA Learning Problem Setup

- **H1**: Moving average
- **H2**: Frequency weighted ratio, threshold = \( V \)
- **H3**: Frequency weighted ratio, threshold = \( A \)
- **Q1**: Opposite role
- **Q2**: Same role
- **T1**: Total
- **T2**: Since last trade
- **U**: Number of trades left
- **V**: Value of next unit to be traded

**Actions**
- \( A \): Offset from \( V \)

**Rewards**
- \( R \): Difference between unit valuation and trade price

**State Space**
- History of recent trades
- Quotes
- Time
- Pending Trades
### EGTA/RL Round 1

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Payoff</th>
<th>NE</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan ZI Zlb tq</td>
<td>248.1</td>
<td>1.000 ZI</td>
<td>L1</td>
</tr>
<tr>
<td>L1</td>
<td>242.5</td>
<td>1.000 L1</td>
<td></td>
</tr>
</tbody>
</table>

### EGTA/RL Round 2

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Payoff</th>
<th>NE</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan ZI Zlb tq</td>
<td>248.1</td>
<td>1.000 ZI</td>
<td>L1</td>
</tr>
<tr>
<td>L1</td>
<td>242.5</td>
<td>1.000 L1</td>
<td></td>
</tr>
<tr>
<td>ZIP</td>
<td>248.0</td>
<td>1.000 ZIP</td>
<td></td>
</tr>
<tr>
<td>GD</td>
<td>248.6</td>
<td>1.000 GD</td>
<td>L2-L8 L9</td>
</tr>
<tr>
<td>L9</td>
<td>246.1</td>
<td>0.531 GD 0.469 L9</td>
<td>L10</td>
</tr>
</tbody>
</table>
**EGTA/RL Rounds 3+**

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Payoff</th>
<th>NE</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strategy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>L10</td>
<td>248.0</td>
<td>0.191 GD</td>
<td>L10</td>
</tr>
<tr>
<td>L11</td>
<td>246.2</td>
<td>1.000 L11</td>
<td>L11</td>
</tr>
<tr>
<td>GDX</td>
<td>245.8</td>
<td>0.192 GDX</td>
<td>L11</td>
</tr>
<tr>
<td>L12</td>
<td>245.8</td>
<td>0.049 L11</td>
<td>L12</td>
</tr>
<tr>
<td>L13</td>
<td>245.6</td>
<td>0.872 L12</td>
<td>L13</td>
</tr>
<tr>
<td>RB</td>
<td>245.6</td>
<td>0.872 L12</td>
<td>L13</td>
</tr>
</tbody>
</table>

**Final champion**

---

**Strategy Exploration Problem**

- **Premise:**
  - Limited ability to cover profile space
  - Expectation to reasonably evaluate all considered strategies

- Need deliberate policy to decide which strategies to introduce

- **RL for strategy exploration**
  - Attempt at best response to current equilibrium
  - Is this a good heuristic (even assuming ideal BR calc?)
Example

Introduce strategies in order: A1, A2, A3, A4

Regret may increase over subsequent steps!

<table>
<thead>
<tr>
<th>Strategy Set</th>
<th>Candidate Eq.</th>
<th>Regret wrt True Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A1}</td>
<td>(A1,A1)</td>
<td>3</td>
</tr>
<tr>
<td>{A1,A2}</td>
<td>(A2,A2)</td>
<td>4</td>
</tr>
<tr>
<td>{A1,A2,A3}</td>
<td>(A3,A3)</td>
<td>5</td>
</tr>
<tr>
<td>{A1,A2,A3,A4}</td>
<td>(A4,A4)</td>
<td>0</td>
</tr>
</tbody>
</table>

Regret Surface

FPSB2 Regret Surface
Exploration Policies

- **RND**: Random (uniform) selection
- **Deviation-Based**
  - **DEV**: Uniform among strategies that deviate from current equilibrium
  - **BR**: Best response to current equilibrium
  - **BR+DEV**: Alternate on successive iterations
  - **ST(τ)**: Softmax selection among deviators, proportional to gain
- **MEMT**:
  - Select strategy that maximizes the gain (regret) from deviating to a strategy outside the set from *any* mixture over the set.

### CDA ↓ 4

![Graph showing expected regret over steps for different exploration policies](image-url)

- **MEMT**
- **DEV**
- **RND**
- **BR**
- **ST**: Various softmax selection strategies with different parameters.
EGTA Applications

• Market games
  — TAC: Travel, Supply Chain, Ad Auction
  — Canonical auctions: SAAs, CDAs
  — Equity premium in financial trading
• Networking games
  — privacy attacks, routing, wireless access point selection
• Mechanism design

Conclusion: EGTA Methodology

• Extends scope of GT to procedurally defined scenarios
• Embraces statistical underpinnings of strategic reasoning
• Search process:
  — GT for establishing salient strategic context
  — Strategy exploration:
    • e.g., RL to search for best response to that context
  — Principled approach to evaluate complex strategy spaces
• Growing toolbox of EGTA techniques