BEST: A Benchmarked Experiential System for Training

Kevin Gildea, Georgiy Levchuk, Jared Freeman, Aptima Inc.
Shruti Narakesari, Wayne Shebilske, WSU

Presented at 12th CCRTS
Date: June, 21, 2007
How do we recognize **expertise**?
  - Performance

How do we know it is **high**?
  - Benchmarking

How can we **achieve** high expertise?
  - Training
- AF Dynamic Targeting Cell
  - Find targets
  - Classify/ID targets
  - Coordinate assets
  - Approve strike packages/execute
The Complication

- **What if the task is ill-structured or complex?**
  - Can make **mistakes in recognizing achieved expertise**
  - Do not know **impact of training**

- **Why?**
  - Multiple types of expertise combine to affect performance
  - Incorrect/biased observer ratings or measures
  - Weak knowledge of expertise dynamics
What Happens to Training?

- You cannot train what you cannot measure
  - How do you give training if you do not know achieved expertise and impact the training will have?
- Different teams = different learning curves
State of Training Systems

- ITTS systems rarely model team learning, typically model
  - Expert team members (Miller, et al., 2000)
  - Coaches for individuals (Freeman, et al., 2005)
- Qualitative I/O models of team learning are ... qualitative so cannot drive training sims
- Quantitative I/O models of team learning do not drive training sims (Kozlowski, et al., 2001)
Our Contributions

- Definition of expertise & dynamics

- **Assessment:** benchmarking via optimal mission execution solution

- **Improvement:** intelligent training system
  - *Optimization in instructional strategy:* Train the team with expert-selected, annotated, and animated near optimal solutions delivered in feedback
  - *Optimization of instructional strategy:* Select scenarios maximizing likelihood of advancing the team most directly to goal expertise using a POMDP model
Expertise (1-2 vectors, with 1=inadequate; 2= adequate):
- ISR maintaining low risk
- ISR maintaining high coverage
- ISR nominating and DTC designating TSTs
- DTC prioritizing TSTs
- DTC coordinating a strike package plan

Mission:
- # Time sensitive targets
- # Enemy defensive threats
Contribution 2: Assessment

- Scenario Benchmark: Near Optimal Solution Model
  - Agent Model

- Human Performance
  - Analysis of DDD log files
  - “Playback” Model
  - Analysis of chat communications
Contribution 3: Strategy Improvement

- Training = Planning under uncertainty
  - **State** = discrete expertise
  - **Learning dynamics** = impact of training
  - **Control** = training scenarios
  - **State observations** = measures of performance

Partially Observable Markov Decision Processes

Partially Observable Markov Decision Processes

- **Expertise State**
  - hidden
- **Training Scenario**
  - controlled
    - effect on expertise
- **Measures**
  - observed
- **Hidden state transition model:** how the training affects team expertise
  - Variables: $\Pr(\text{current state} | \text{next state, training scenario})$

- **Observation model:** how current scenario and team expertise affect what measures can be observed
  - Variables: $\Pr(\text{observed measure} | \text{next state, training scenario})$
## Comparison to Other Markov Models

<table>
<thead>
<tr>
<th>Markov Models</th>
<th>Do we have control over the state transitions?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>Markov Chain</td>
<td>YES</td>
</tr>
<tr>
<td>Markov Decision Process</td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>HMM</td>
<td>NO</td>
</tr>
<tr>
<td>Hidden Markov Model</td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>POMDP</td>
<td>YES</td>
</tr>
<tr>
<td>Partially Observable Markov Decision Process</td>
<td></td>
</tr>
</tbody>
</table>
How Does POMDP Work?

- Benefit of training defined using “reward” of visiting expertise state nodes
  - Want to maximize the expected total reward of training

- Can incorporate specific “training path” restrictions
  - Depends on trainer’s objectives

\[
\text{objective} = \max E \left[ \sum_{t=1}^{\infty} \gamma^t r_t \right]
\]
POMDP solution is provided as a "policy graph" = look-up table

- Policy nodes = abstractions for belief about expertise state
  - Policy node is described using (PolicyNodeID, StepID)
  - pair

- Each node specifies training scenario to be given to the team

- For each feasible observation – specify next policy node

---

### How do Trainers Use POMDP?

- Policy Graph → Lookup Table

POMDP Policy Solution

<table>
<thead>
<tr>
<th>StepID</th>
<th>Policy Node ID</th>
<th>Scenario ID</th>
<th>Scenario Name</th>
<th>Policy Node ID if Observation =</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>11111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>X</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>14</td>
<td>X</td>
<td>14</td>
</tr>
</tbody>
</table>

- 0. Start
- 1. Use this scenario
- 2. Receive Observation
- 3. Mark next policy node

---

How do Trainers Use POMDP?

- Policy Graph → Lookup Table

POMDP solution is provided as a "policy graph" = look-up table

- Policy nodes = abstractions for belief about expertise state
  - Policy node is described using (PolicyNodeID, StepID)
  - pair

- Each node specifies training scenario to be given to the team

- For each feasible observation – specify next policy node
Empirical Validation: Task & Team

- **Find:** *Detect and differentiate (T.O.1)* Time Sensitive Targets (TSTs) ISR
  - Minimize risk to assets
  - Maintain high coverage of targets
  - Nominate & designate TSTs

- **Fix:** Complex communications to detect and differentiate (T.O.1) & *prioritize (T.O.2)* TSTs

- **Target:** Complex communications to *coordinate attack assets (T.O.3)*

- Approve strike package

- **Team:** 7 Ss
  - Participants: ISR, DTC Chief, Ground Track Coordinator (GTC), Attack Coordinator (AC), Target Duty Officer (TDO)
  - Confederates: Senior Offensive Duty Officer (SODO), Chief of Combat Operations (CCO)
Empirical Validation: Procedure

- Participants: 7 undergraduates
- Phases
  - Phase I: Declarative & procedural training + practice (50 hrs)
  - Phase II: 49 trials + feedback (49 hrs)
    - Consistent enemy
    - Scale-up targets & threats
  - Phase III: 18 trials (18 hrs)
    - Inconsistent enemy
    - Scale-up targets & threats
- Measure
  - Quality of the proposed strike package (TO3) for each TST, determined by expert ratings (96% agreement)
- Design (within Ss)
  - Protocol (POMDP vs. control)
  - Phase (II vs. III)
  - Test (pre vs. post)
  - (Counterbalanced scenarios & order)
Results

Phase II training produces mid/high competency operational team

Near transfer

Far transfer

1. Teams can learn the task (p<.01)
2. Far transfer degrades performance (p<.01)
3. Controls learn slowly if at all (p>.05)
4. POMDP condition learns rapidly (p<.01)

- Phase II training produces mid/high competency operational team
Aptima is leading research in optimizing instructional strategy

- Optimization *in* instruction: near-optimal solutions as feedback
- Optimization *of* instruction: POMDP model driven scenario selection

Our developed automated intelligent tutoring system significantly outperformed expert-based training solutions