Learning the NRL Navigation Task

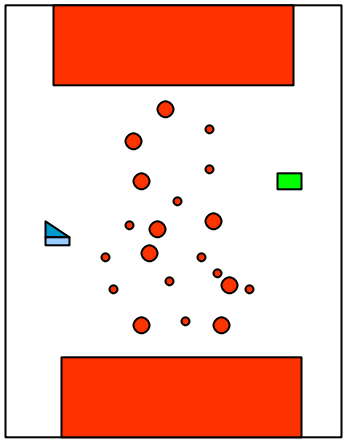
Devika Subramanian
Rice University

joint work with:
Diana Gordon, NRL and Sandra Marshall, SDSU

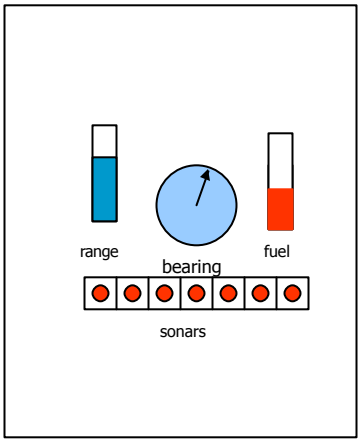
Talk outline

- The Navigation task: a challenge for human learning
- Features of human performance
- Results of cognitive modeling
- Machine learning the task
- Challenges to machine learning

The NRL Navigation Task



Global View

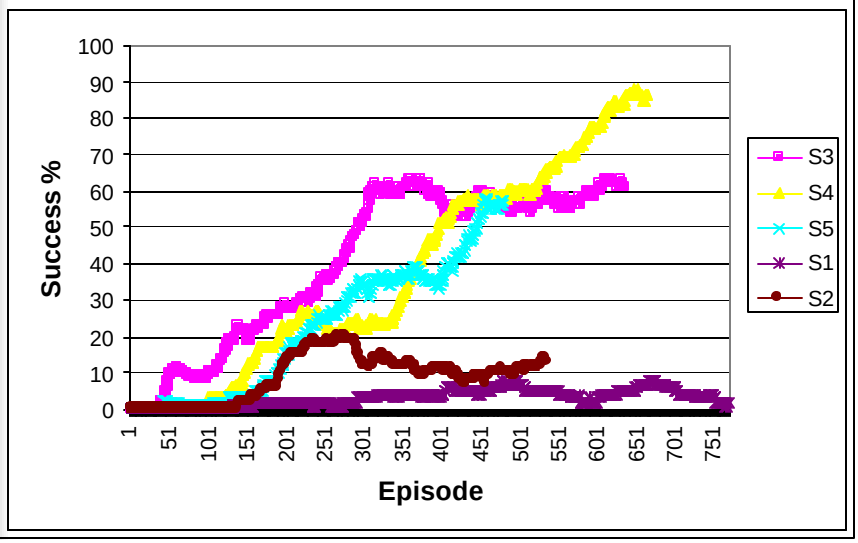


Instrument View

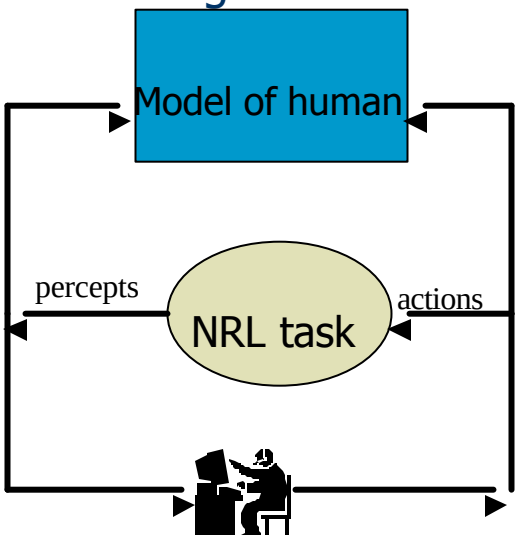
Challenges for a human learner

- Need for competent visual-motor coordination.
- Need for rapid decision making.
- Limited intermediate feedback.
- To achieve success, subject must learn to coordinate solutions to getting to target and avoiding mines.

Learning curves (success)



The modeling task



Track the evolution of the action policy
f: percepts → actions

Cognitive modeling by machine learning

Find approximation g of the subject's policy function f from samples of f (i.e., the execution traces).

Model evaluation criteria

- Fit to learning curves.
  - f and g are equivalent if they generate the same learning curve.
- Fit to action probability distributions.
  - f and g are equivalent if the distribution of actions for specific classes of situations they generate are close-enough.
- Fit to sequences of motor actions.
  - f and g are equivalent if they generate the same sequences of actions.

Modeling deviation from optimal strategy

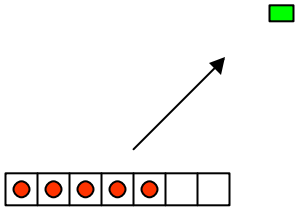
- Compare human performance to optimal solution to task.
  - Instead of describing what subject is learning, track the evolution of the difference between subject's policy and the optimal policy.
- Advantage
  - deviations from optimal can be the basis for directed training of subjects.
- Disadvantage
  - humans may not adopt anything close to the conceptualization needed for optimal play.

A near-optimal player

- A three-part deterministic controller solves the task!
- The only information required about the previous state is the last-turn made.
- A very coarse discretization of the state space is needed: about 1000 states!
- Discovering this solution was not easy!

Part 1: Seek Goal

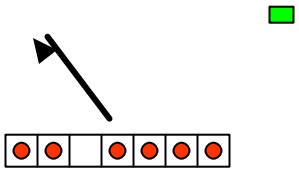
There is a clear sonar in the direction of the goal.



If the sonar in the direction of the goal is clear, follow it at speed of 20, unless goal is straight ahead, then travel at speed 40.

Part 2: Avoid Mine

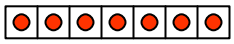
There is a clear sonar but not in the direction of the goal.



Turn at zero speed to orient with the first clear sonar counted from the middle outward. If middle sonar is clear, move forward with speed 20.

Part 3: Gap Finder

There are no clear sonars.



If the last turn was non-zero, turn again by the same amount, else initiate a soft turn by summing the right and left sonars and turning in the direction of the lower sum.

Learning trends in data

- Subjects have relatively static periods of action policy choice punctuated by radical shifts.
- Sequence of events associated with a shift:
  - altered conceptualization of the task.
  - change in perception strategy.
  - change in action strategy.
  - significant performance improvement.

What are successful subjects learning?

- To follow bearing better in states in part 1.
- To slow down significantly when turning.
- To turn minimally to avoid mines in states in part 2.
- To turn in place consistently to find gaps in states in part 3.

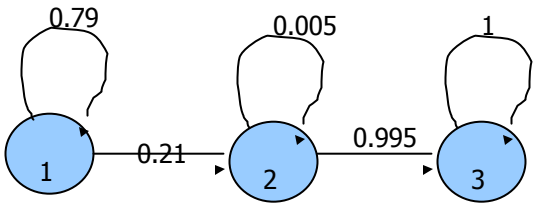
Subject 5, sonars on left blocked

- Subject learns to go faster straight past the mines to the left. Also learns to veer right a bit.
- Learns to turn at zero speeds.

Evolution of gap finding strategy

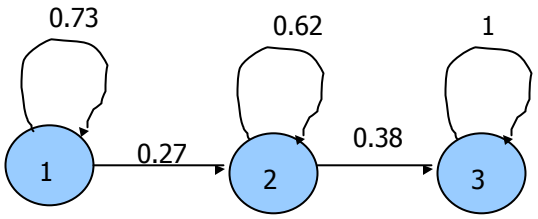
- Subject 5: episodes 45-67 and episodes 68-90 on day 2.
- Subject learns to turn in place.
- HMM models for gap finding action sequences.

Pre-shift gap finding strategy



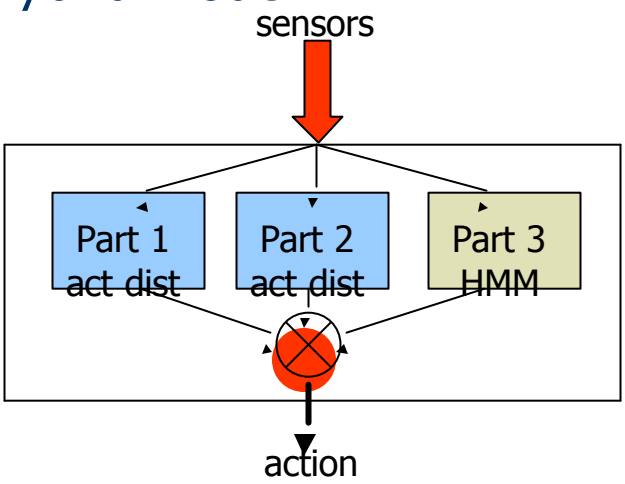
0	**0**	0.84	0.00	0.36
right	**0**	0.03	0.003	0.33
left	**0**	0.05	0.00	0.31
other		0.08	0.997	0.00

Post-shift gap finding strategy



0	**0**	0.82	0.05	0.37
right	**0**	0.024	0.00	0.553
left	**0**	0.025	0.95	0.07
other		0.131	0.00	0.00

A hybrid model



A very small number of parameters is sufficient to capture subject. Can acquire subject model online!

Results

Pre-shift	Successes	Explosions	Timeouts	Total episodes
S5	0	12	11	23
Model	0	17	6	23
Post-shift	Successes	Explosions	Timeouts	Total episodes
S5	0	2	13	15
Model	0	4	11	15

A better fit than using C4.5.

Machine learning of NRL task

- What does it take to get machines to learn task?
- Can machine learners achieve higher levels of competence?
- How does the sample complexity of machine learning compare with humans?
- Can we use machine learning to improve human learning?

Challenges for a machine learner

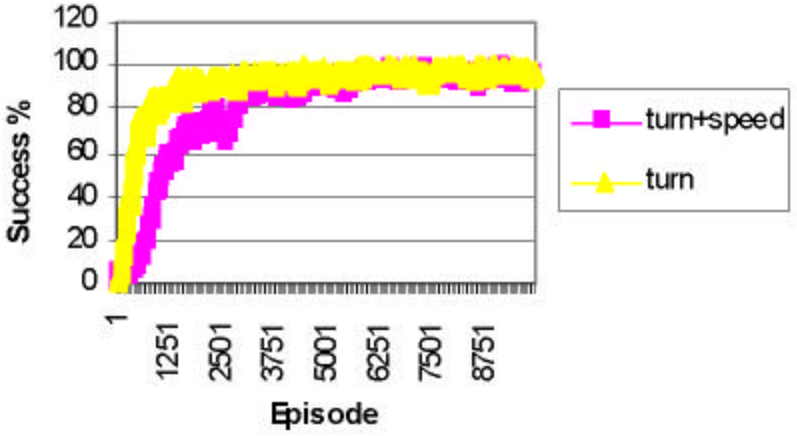
- Enormous, irregular state space: 10^{15} states.
- Large action space: 153 actions/state.
- Lack of intermediate feedback: binary feedback at the end of a long sequence of moves.

The progress function

r(s,a,s') = 0 if s' is a state where player hits mine.
= 1 if s' is a goal state
= 0.5 if s' is a timeout state

= 0.75 if s is a Part3 state and s' is a Part1 or Part2 state
= 0.5 + ✍sum of sonars/1000 if s' is a Part3 state
= 0.5 + ✍range/1000 + abs(bearing - 6)/40 otherwise

Results of learning complete policy



Lessons from machine learning

- Why task is hard: most frequently occurring state occurs 45% of time, all others are less than 5%.
- Long sequence of moves makes credit assignment hard.
- Staged learning makes task easier; and might help humans acquire task easier.
- Need for a locally non-deceptive reward function to speed up training. Can giving progress function as hints to human players help?

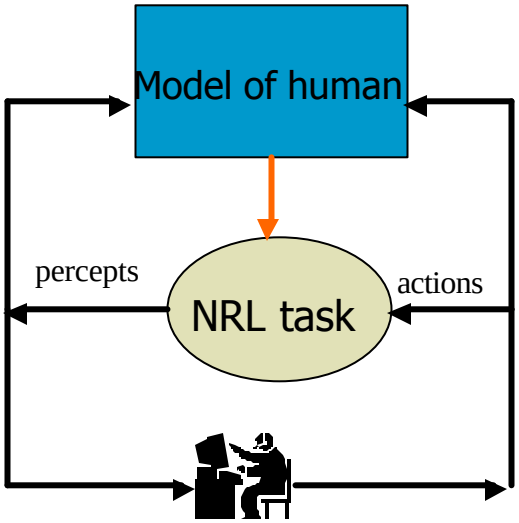
Conclusions

- We have used inductive machine learning techniques to construct compact cognitive models from the vast empirical visual-motor data gathered from subjects.
- Cognitive modeling poses new problems in the design of automated discretization techniques for inductive machine learning.

Conclusions

- We have studied machine learners for the task and used the results to understand complexities of task as well as to suggest new staged protocols for training humans.
- Machine learning the NRL task has pushed the science and engineering of reinforcement learning.

Closing the modeling loop



Track human learning online and tailor task to speed up acquisition.