A POMDP Approach to Antenna Selection Problem in MIMO Systems

Sinchu P

20 Oct., 2012

Outline of the Presentation

- An Overview of POMDP
- Antenna Selection
- System model
- POMDP Formulation
- Simulation results
- Future Work

- Robot navigation in a building

- Robot navigation in a building
- Available information
  - Map of the building
  - Observation
  - Deterministic actions: eg. North South East West

An Overview of POMDP

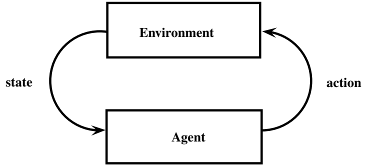
- Robot navigation in a building
- Available information
  - Map of the building
  - Observation
  - Deterministic actions: eg. North South East West
- Goal: Reach nth room in the mth floor
- Penalty for being in “non-goal” locations (optional)

An Overview of POMDP

- Robot navigation in a building
- Available information
  - Map of the building
  - Observation
  - Deterministic actions: eg. North South East West
- Goal: Reach nth room in the mth floor
- Penalty for being in "non-goal" locations (optional)
- Ideal case: Markov property holds
- Practical scenario: Markov property doesn't hold directly

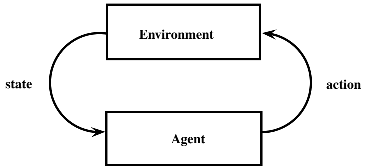
An Overview of POMDP

- MDP: Given a complete and correct model of the world dynamics and a reward structure, find an optimal way to behave.



An Overview of POMDP

- MDP: Given a complete and correct model of the world dynamics and a reward structure, find an optimal way to behave.



- Finite set of states $\mathcal{S}$
- Finite set of actions $\mathcal{A}$
- State transition function $T(s, a, s')$
- Rewards function $R(s, a)$

An Overview of POMDP

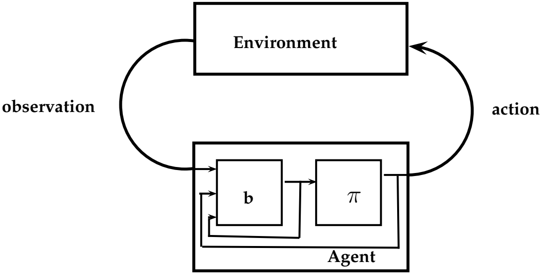
- Policy:$\pi_t(s)$ is a situation-action mapping.
- Value function:
  $V_{\pi,t}(s) = R(s, \pi_t(s)) + \beta \sum_{s' \in \mathcal{S}} T(s, \pi_t(s), s') V_{\pi,t-1}(s')$
  (Bellman Equation)
- For infinite horizon discounted case, there exists a stationary policy Π^* that is optimal for every state. [Howard]

An Overview of POMDP

- Policy:$\pi_t(s)$ is a situation-action mapping.
- Value function:
  $V_{\pi,t}(s) = R(s, \pi_t(s)) + \beta \sum_{s' \in \mathcal{S}} T(s, \pi_t(s), s') V_{\pi,t-1}(s')$
  (Bellman Equation)
- For infinite horizon discounted case, there exists a stationary policy Π^* that is optimal for every state. [Howard]
- Value iteration for finding optimal policy.
- Algorithm:

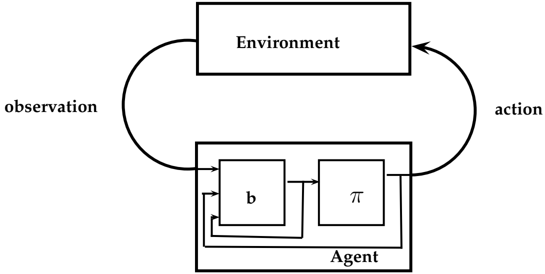
  $V_1(s) = 0$ for all s
  loop $t = 1$
    $t = t + 1$
    loop for all $s \in \mathcal{S}$
      loop for all $a \in \mathcal{A}$
        $Q_t^a(s) = R(s, a) + \beta \sum_{s' \in \mathcal{S}} T(s, a, s') V_{t-1}(s')$
      end loop
    $V_t(s) = max_a Q_t^a(s)$
    end loop
  until $|V_t(s) - V_t - 1(s)| < \epsilon$

- POMDP:



An Overview of POMDP

- POMDP:



- Finite set of states $\mathcal{S}$
- Finite set of actions $\mathcal{A}$
- Finite set of observations Ω
- State transition function $T(s, a, s')$
- Rewards function $R(s, a)$
- Observation function $O(s', a, o)$

An Overview of POMDP

- Belief state; A probability distribution over the state of the world
  :comprise a sufficient statistics for the past history, the process over belief states is Markov [Sondik].

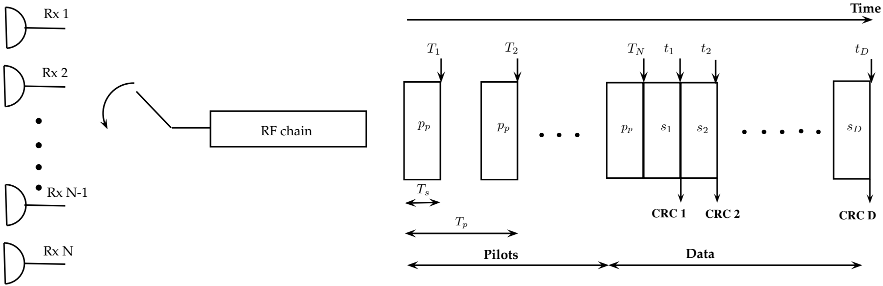
$$b'(s') = P(s'|o, a, \mathbf{b})$$
$$= \frac{O(s', a, o) \sum_{s \in \mathcal{S}} T(s, a, s') b(s)}{P(o|a, \mathbf{b})}$$

- Value function
  $V_p(s) = R(s, a) + \beta \sum_{s' \in \mathcal{S}} T(s, a(p), s') \sum_{o_i \in \Omega} O(s', a(p), o_i) V_{o_i(p)}(s')$
  $V_p(b) = \sum_{s \in \mathcal{S}} b(s) V_p(s)$
- finite horizon: piece-wise linear and convex
- infinite horizon: continuous and convex

Antenna Selection

- *Motivation*: Expensive RF components and relatively inexpensive antenna elements.
- *Technique*: Adaptively switch a smaller number of analog chains to a subset of available antennas such that the diversity of the full complex system is retained.
- *Fact*: Diversity order with perfect CSI is achievable with imperfect CSI.

System Model



- Rayleigh fading channel with Jake's Doppler spectrum
- Single transmit antenna
- Multiple receive antennas with single RF chain
- CSI estimated from pilots
- Non-uniformly delayed channel estimates
- CRC at the end of every symbol

POMDP formulation for SEP minimization

- States: Discretized channel states ($N^{|\mathcal{S}|}$)
- Actions: Select one of the N antennas at the start of every symbol
- Transition probabilities: Obtained for a given value of normalized Doppler
- Observation: CRC bits
- Reward: Probability of correctly decoding the symbol
- Horizon: finite or infinite

  Optimum policy found using POMDP solver tool

POMDP formulation for SEP minimization:Finite Horizon

- N=4, $|\mathcal{S}|$=2 (Gibert-Elliot Model)

Pilot symbol received on k^{th} antenna

$$
\begin{aligned}
y &= |h|e^{j\phi}p + w \qquad w \sim \mathcal{CN}(0, N_0) \\
y' &= |h|e^{j\phi} + w' \qquad w' \sim \mathcal{CN}(0, 1/snr) \\
y'' &= \sqrt{\gamma}e^{j\phi} + w'' \qquad w'' \sim \mathcal{CN}(0, 1), \quad \gamma \triangleq \| h|^2 snr
\end{aligned}
$$

An unbiased estimate for γ,

$$\hat{\gamma} = |y''|^2 - 1$$

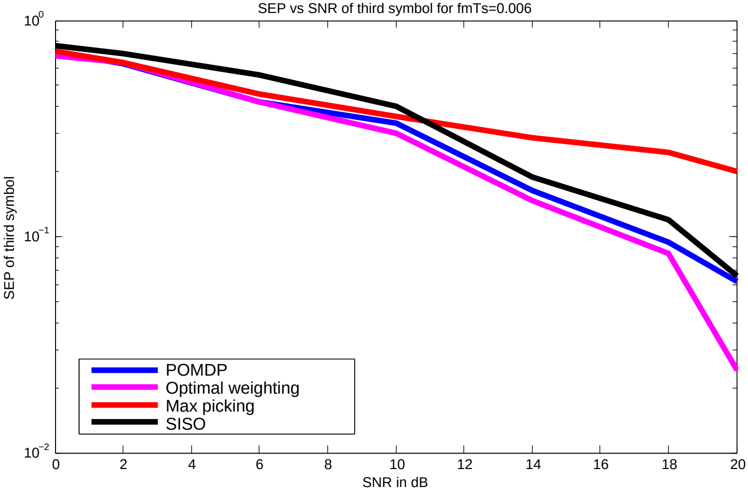
POMDP formulation for SEP minimization: Finite Horizon

- To find initial belief state : $P(\gamma_a < \gamma < \gamma_{a+1} | \hat{\gamma})$

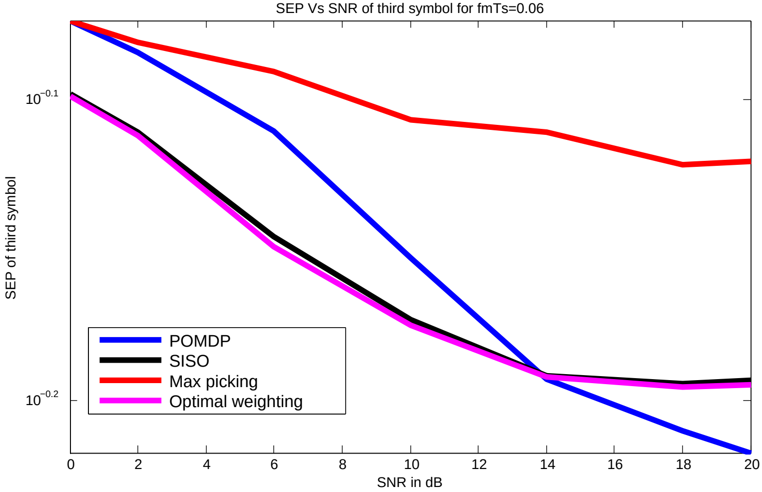
$$= \frac{P(\hat{\gamma} | \gamma_a < \gamma < \gamma_{a+1}) P(\gamma_a < \gamma < \gamma_{a+1})}{P(\hat{\gamma})}$$

- Find the optimal action given by POMDP solution
- Decode symbols using CSI
- Update belief state on receiving each CRC bit

Results


SEP vs SNR of third symbol for fmTs=0.006
SEP of third symbol
SNR in dB
10^0
10^-1
10^-2
0
2
4
6
8
10
12
14
16
18
20
POMDP
Optimal weighting
Max picking
SISO

Results


SEP Vs SNR of third symbol for fmTs=0.06
SEP of third symbol
$10^{-0.1}$
$10^{-0.2}$
0
2
4
6
8
10
12
14
16
18
20
SNR in dB
POMDP
SISO
Max picking
Optimal weighting

Future Work

- Effect of number of channel states
- Infinite horizon model
- Structural results of the optimal policy