Partially Observable Markov Decision Processes

Two aspects of intrinsic complexity:
- super-exponential growth of the number of vectors
- identification of all useful vectors

ONLINE APPROXIMATION ALGORITHMS

Although they usually need a value function computed offline for the leaves of the tree, these methods are classified as 'pure' online algorithms. Also note that these methods are by definition approximation algorithms.

Given an initial belief state, only a finite number of belief states are reachable. This leads to an algorithm for selecting actions that is singly exponential in the horizon allowed for by the online planning time.

TREE-SEARCH ALGORITHMS

These algorithms can be conceptualized by viewing a POMDP as a game between the agent and the environment. The agent has to select actions at each belief node and then the environment randomly selects the next observation. The best action is obtained from a search to a given depth.

- BRANCH AND BOUND PRUNING
  - These algorithms maintain upper and lower bounds on each node of the tree, and these are updated recursively up the tree, unwinding from the fringe nodes.
  - general equations
  - RTBSS
    - Paquet, Tobin, & Chaib-draa, 2005
    - Perform a d-step lookahead and use U(b) at the bottom of the tree.
    - equations
- MONTE CARLO SAMPLING

  Used to lower the complexity of exploring the whole observation space. Could also be used to reduce complexity of the action space - HOW?
  - MCALLESTER AND SINGH
    - McAllester & Singh, 1999
    - Expands on the work on MDPs by Kearns et al, 1999 and Boyen & Koller, 1998 and provides two analyses of belief state simplifiers:

      - The first result states that under KL-epsilon belief state simplification and for eta-mixing POMDPs, a near-optimal policy in the SIMPLIFIED belief state MDP performs nearly optimal in the TRUE belief state MDP.

      - The second result states that under L_1-epsilon belief state simplification and any POMDP, a near optimal policy is near optimal for the TRUE belief state only at the beginning (ie., t is present in the bound equation).
    - equations
  - ROLLOUT
    - Bertsekas & Castanon, 1999
  - PARALLEL ROLLOUT
    - Chang, Givan & Chong, 2004
- HEURISTIC SEARCH
  - general equations
  - SATIA AND LAVE
    - Satia & Lave, 1973
    - equations
  - BI-POMDP
    - Washington, 1997
    - Guides the search toward nodes that maximixe the upper bound. Does not take into account the observation model nor the discount factor.
    - equations
  - AEMS
    - Ross and Chaib-draa, 2007
    - Combination of Satia and Lave and BI-POMDP here \pi_t represents the probability that action a is optimal in its parent belief b given the bounds in tree T
    - equations

ITERATIVE IMPROVEMENTS TO THE VALUE FUNCTION

- ITERATIVE PERSEUS
  - Shani et al, 2005
  - Adaptation of the original PERSEUS algorithm. It uses two value functions, the current one V and the next one V'. V' is modified untill it remains stable for a period of time. When this happens, V' becomes the active function V.

EXACT ALGORITHMS

Two main approaches for computing useful linear functions:
- generate and test:
  enumerate all possible linear functions first, then test for usefulness and prune
  there are refinements to this method that do early pruning
- search and generate:
  compute a useful linear function for a single belief point:
  the problem here is to locate all belief points that are relevant

Two aspects of intrinsic complexity:
- super-exponential growth of the number of vectors
- identification of all useful vectors

- VALUE ITERATION
  - Early POMDP solutions performed exact value iteration using PWLC value functions.
  - Bellman, 1957
  - Sondik, 1971
  - Monahan, 1982
  - equations
- POLICY ITERATION — Howard, 1960; Sondik, 1978
- WITNESS — Kaelbling, Littman & Cassandra, 1999
- STOCHASTIC DECISION TREES
  - Raiffa, 1970
  - This algorithm assumes a fixed initial belief state, and seems to only make sense in the finite-horizon case
- INCREMENTAL PRUNING — Cassandra et al., 1997

OFFLINE APPROXIMATION ALGORITHMS

VALUE FUNCTION APPROXIMATIONS

- FULLY OBSERVABLE MDP APPROXIMATIONS
  - MDP POLICY HEURISTICS
    - MAXIMUM LIKELIHOOD
      - Nourbakhsh et al., 1995
      - Select the most likely state according to b(s) and act using MDP.
    - VOTING
      - Simmons & Koenig, 1995
      - Weighted vote for each state's best action.
    - DUAL-MODE CONTROLLER
      - Cassandra et al., 1996
      - Keeps track of the belief's entropy and tries to reduce its expected value.
  - MDP
    - Take an optimistic stance: assume the world is fully observable.
    - Astrom, 1965; Lovejoy, 1993
    - equations
  - QMDP
    - Push fully observability one step away: assume all uncertainty will disappear after the first action.
    - Littman, Cassandra & Kaelbling, 1995
    - equations
- UNOBSERVABLE MDP APPROXIMATIONS
  - The problem of finding the optimal solution for the finite-horizon UMDP is NP-hard, and the infinite-horizon case is undecidable. This method provides a lower bound to the exact update.
  - equations
- FIB
  - Hauskrecht, 2000
  - The Fast Informed method takes some of the partial observability into account.
    How?
    The value function obtained via this method is piecewise linear and consists of at most
    |A| different linear functions, each corresponding to one action.
    Both the finite-horizon solution and inifinite-horizon fixed point solution to this method
    are computable efficiently.
    The main idea is to select the best linear function for every observation and every current
    state SEPARATELY. This differs from the exact update where we seek a linear function that
    gives the best result for every observation and the COMBINATION of all states.
    - EXTENSIONS OF FIB
  - equations
- POINT-BASED APPROXIMATIONS

  These methods update entire vectors on a given set of belief points. They can be classified into batch update and asynchronous update methods.
  - POINT-BASED DYNAMIC PROGRAMMING — Cheng, 1988
  - INCREMENTAL LINEAR-FUNCTION APPROACH — Hauskretch, 2000
  - ZHANG'S PBVI — Zhang & Zhang, 2001
  - PINEAU'S PBVI

    Perhaps the most widely used PBVI algorithm.
    - Pineau et al., 2003, 2006
    - Pineau & Gordon, 2005
  - PERSEUS
    - Spaan & Vlassis, 2005
    - randomized point-based value iteration
      for this algorithm there is no clear notion of a horizon
      it makes sense to apply it to the infinite-horizon case
    - equations
  - HEURISTIC SEARCH
    - HSVI — Smith & Simmons, 2004
    - RTDP-BEL? — Geffner & Bonet, 1998
    - BRTDP
      - McMahan et al., 2005
      - Maintains an upper and lower bound on V* and updated those belief points where the difference between the two is large.
  - PBUA — Poon, 2001
- GRID-BASED APPROXIMATIONS WITH INTERPOLATION-EXTRAPOLATION
  - updates are computed for only a finite set of grid points G
  - equations
  - FIXED GRID
    - CONVEX FAMILY
      - Averagers by Gordon, 1995 are a more general case.
        This family includes approaches such as:
        - nearest neighbor
        - kernel regression
        - linear point interpolations
      - GRID-BASED MDP

        Using a grid method can be stated as finding the sequence of values for all grid points in G.
        Some of the convex rules allow a conversion to a grid-based MDP that can then be solved efficiently.
      - equations
  - VARIABLE GRID
- FIXED STRATEGY APPROXIMATIONS
  - the value function consists of |M| linear functions
  - equations
- CURVE FITTING - LEAST SQUARES FIT
  CURVE APPROXIMATIONS

POLICY APPROXIMATION

These methods restrict the space of policies by searching among:
- memoryless policies
- policies based on truncated histories
- memory bounded finite-sate controllers

- REINFORCE — Williams, 1992
- VAPS
  Value and Policy Search — Baird & Moore, 1999
- GPOMDP — Baxter & Bartlett, 2000
- BPI
  Bounded Policy Iteration — Poupart & Boutilier, 2003, 2004

HIERARCHICAL APPROACHES

- HPOMDP — Theocharous, 2002
- POLICY-CONTIGENT ABSTRACTION
  PolCA+ — Pineau, 2004
- HIERARCHICAL FSCs — Hansen & Zhou, 2003