Action commitment Strategy For Situated Planning

Tianyi Gu, Wheeler Ruml

Problem to Address

online planning -> action selection -> when to commit to TLA = A* vs real-time

Assumptions

1. system can’t be uncontrolled -> action duration creates commitment deadline
2. search tree structure = order of decisions is fixed
3. no replanning, decision=commitment to eventual execution
4. assume deterministic system, restart search at new state if necessary
5. only question is when to `reroot the tree’ at a successor of the root = should we ever do this before it’s necessary?
6. we do not propose a expansion strategy, just a commitment strategy (that must take expansion strategy into account at least implicitly)

Previous Work

Mo’RTS (SoCS-15) decides whether to take identity actions

commitment strategy is unprincipled hack

Slo’RTS (ICAPS-17) decides when more thinking time is worthwhile

Problem setting

Commitment queue has all the committed actions

During current search (each expansion), we can choose to:

1) commit now: commit the best top-level action and re-root the search tree
2) commit later: continue the current search (do more expansion)

We are going to choose whichever gives us better utility

If we search long enough, execution will reach root and force commitment

Approach

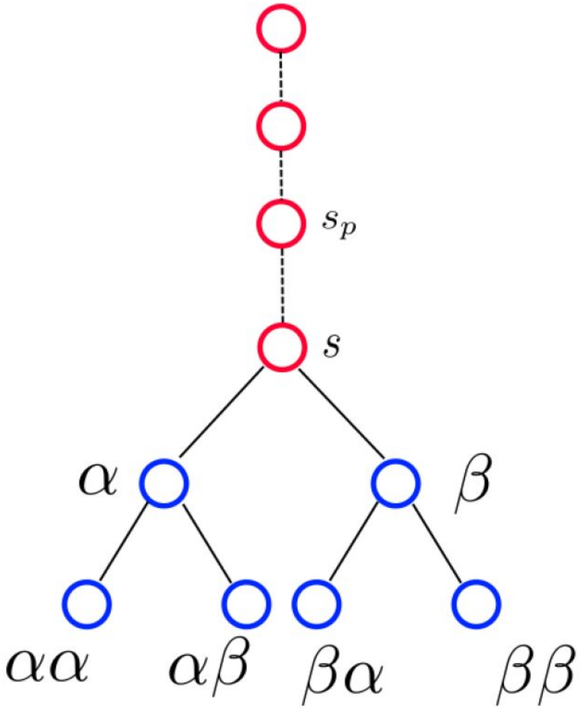
Myopic metareasoning, no multi-step meta-level planning

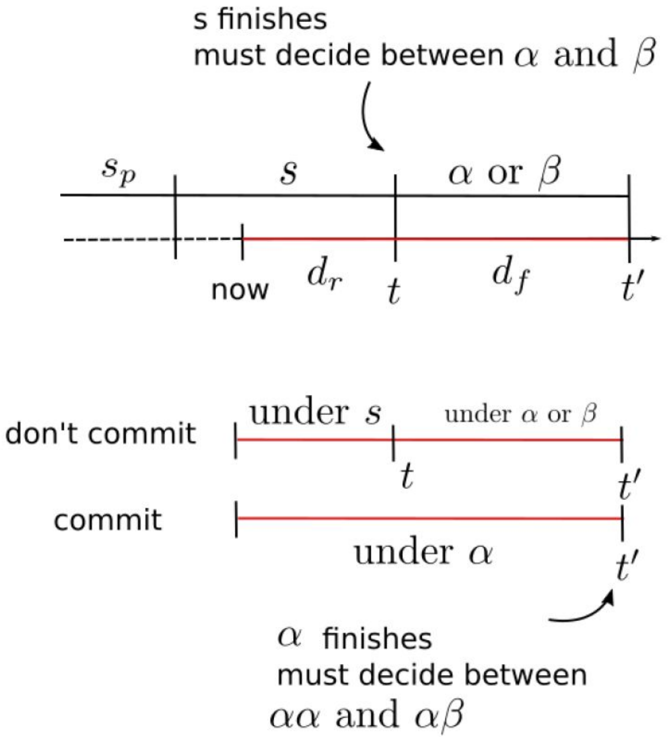
Only consider 2 options:

1. commit now, have more time to consider next action decision
2. don’t commit until deadline, less time for next decision

Estimate utility at deadline for next decision (when we reach children of root)

Problem setting





Effect of doing more search

Doing more search -> f^ could move more

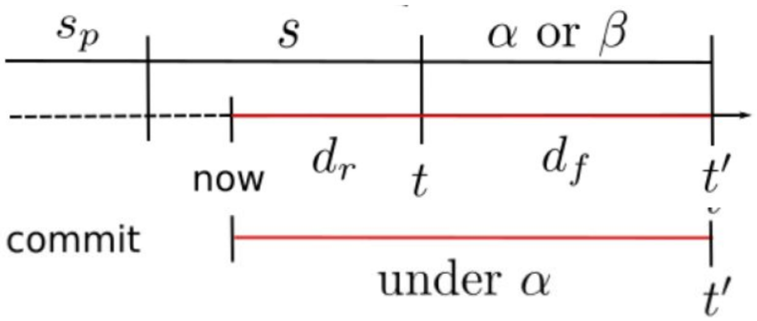
Assumption: the closer to the goal, the higher the variance

0 if little search, f@ belief if all the way to a goal

For every child b of every top-level action a, we model the agent's belief about f^ as :

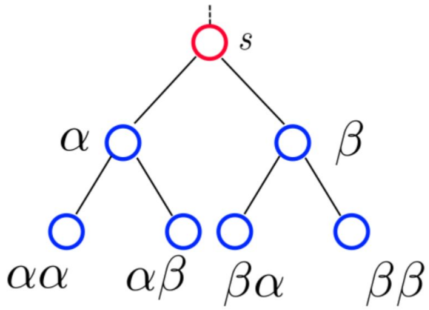
$$p'_{\bar{ab}} \sim \mathcal{N}(\hat{f}(s_{ab}), (\bar{\epsilon}_{b_{ab}} \cdot d(b_{ab}))^2 \cdot \min(1, \frac{d_s}{d(b_{ab})})$$

Utility if we commit **now**



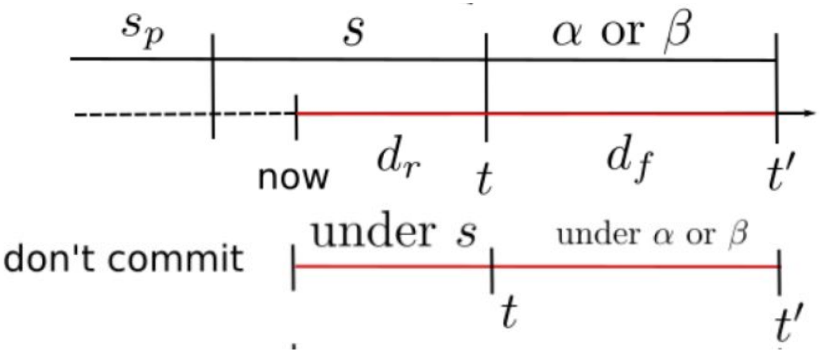
Utility at time t':

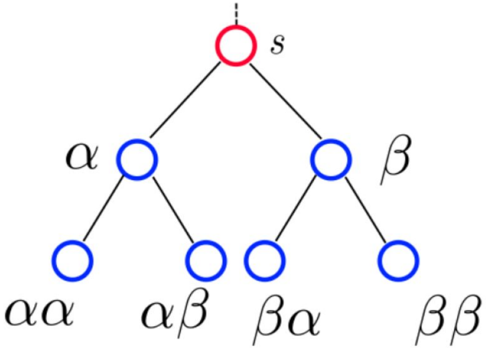
Let $P_s^d(x)$ be the predicted probability of having the belief that $\hat{f}(s) = x$ given d seconds more search under tree node s



$$U_{\text{commit}}^{t'} = \int_{x_{\alpha\alpha}} \int_{x_{\alpha\beta}} P_{\alpha\alpha}^{\frac{d_r+d_f}{2}}(x_{\alpha\alpha}) P_{\alpha\beta}^{\frac{d_r+d_f}{2}}(x_{\alpha\beta}) \min(x_{\alpha\alpha}, x_{\alpha\beta}) dx_{\alpha\alpha} dx_{\alpha\beta}$$

Utility if we commit **later**





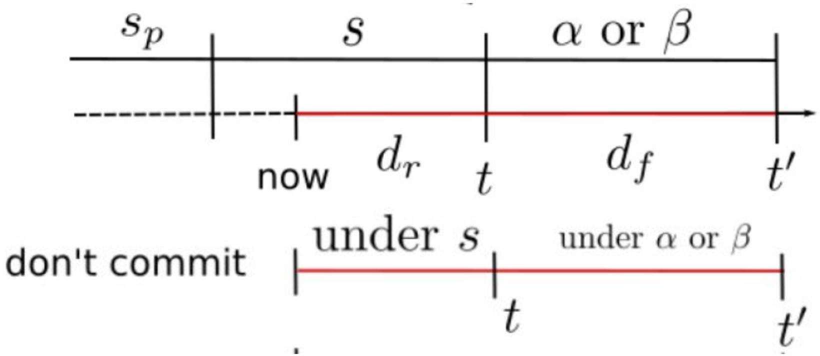
Case 1: after d_r search under α and β, we will believe $\hat{f}(\alpha) \leq \hat{f}(\beta)$

$$U_\alpha^{t'} = \int_{x_{\alpha\alpha}} \int_{x_{\alpha\beta}} P_{\alpha\alpha}^{\frac{\frac{d_r}{2}+d_f}{2}}(x_{\alpha\alpha}) P_{\alpha\beta}^{\frac{\frac{d_r}{2}+d_f}{2}}(x_{\alpha\beta}) \min(x_{\alpha\alpha}, x_{\alpha\beta}) dx_{\alpha\alpha} dx_{\alpha\beta}$$

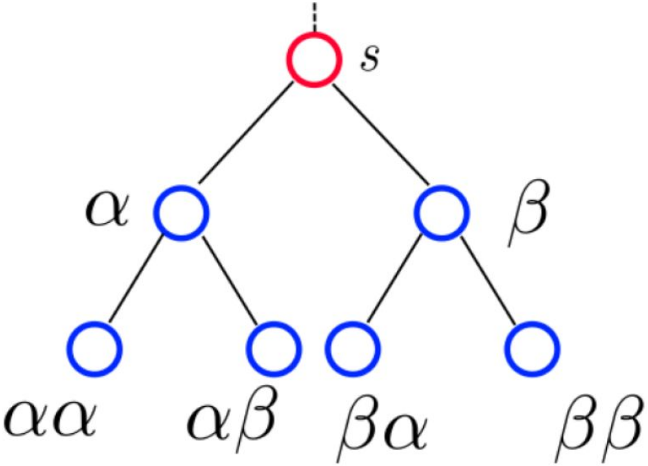
Case 2: after d_r search under α and β, we will believe $\hat{f}(\alpha) > \hat{f}(\beta)$

$$U_\beta^{t'} = \int_{x_{\beta\alpha}} \int_{x_{\beta\beta}} P_{\beta\alpha}^{\frac{\frac{d_r}{2}+d_f}{2}}(x_{\beta\alpha}) P_{\beta\beta}^{\frac{\frac{d_r}{2}+d_f}{2}}(x_{\beta\beta}) \min(x_{\beta\alpha}, x_{\beta\beta}) dx_{\beta\alpha} dx_{\beta\beta}$$

Utility if we commit **later**



Utility at time t' :



$$P_{\text{choose }\alpha} = \int_{-\infty}^{\infty}\int_{x_\alpha}^{\infty} P_\alpha^{\frac{d_r}{2}}(x_\alpha) P_\beta^{\frac{d_r}{2}}(x_\beta)\, dx_\beta\, dx_\alpha$$

$$U^{t'}_{\text{don't commit}} = P_{\text{choose }\alpha} U^{t'}_\alpha + (1 - P_{\text{choose }\alpha}) U^{t'}_\beta$$