Metareasoning for Concurrent Planning and Execution

Dylan O’Ceallaigh and Wheeler Ruml

University of New Hampshire
Department of Computer Science

Grateful thanks to NSF and LAAS-CNRS for support.
When to plan and when to act?
When to plan and when to act?

Assumption: always planning (dedicated core)
When to plan and when to act?

Assumption: always planning (dedicated core)
Question: when to commit?
When to plan and when to act?

Assumption: always planning (dedicated core)

Question: \textit{when to commit?}

1. off-line: complete plan before acting
   implicit \textit{identity action} that preserves state
When to plan and when to act?

Assumption: always planning (dedicated core)

Question: when to commit?

1. off-line: complete plan before acting
   implicit *identity action* that preserves state

2. real-time: plan incrementally, commit to current best
   never execute *identity*, plan while acting
When to plan and when to act?

Assumption: always planning (dedicated core)

Question: when to commit?

1. off-line: complete plan before acting
   implicit *identity* action that preserves state
2. real-time: plan incrementally, commit to current best
   never execute *identity*, plan while acting

How to choose?
Is there a middle ground?
Can we plan/commit dynamically?
How can we decide in a principled way?
When to plan and when to act?

Assumption: always planning (dedicated core)
Question: when to commit?

1. off-line: complete plan before acting
   implicit *identity action* that preserves state
2. real-time: plan incrementally, commit to current best
   never execute *identity*, plan while acting

How to choose?
Is there a middle ground?
Can we plan/commit dynamically?

How can we decide in a principled way?

Metareasoning!
The Problem Setting

1. planning as forward state-space heuristic search
2. minimize goal achievement time (GAT)
   action ‘cost’ = duration
3. access to an inadmissible heuristic $\hat{h} (+ g = \hat{f})$
4. for simplicity: known deterministic world, serial plan
The Problem Setting

1. planning as forward state-space heuristic search
2. minimize goal achievement time (GAT)
   
   action ‘cost’ = duration
3. access to an inadmissible heuristic $\hat{h} ( + g = \hat{f})$
4. for simplicity: known deterministic world, serial plan

Central acting decisions:

1. execute (a) current-best action or (b) identity action?
2. if (a), how many actions?
plan when it appears worthwhile!
When to Plan More?

- Metareasoning
- Estimating Belief
- IMR
- Simple Problems
- Large Problems 1
- Large Problems 2

How Many Actions?

Conclusion

---

plan when it appears worthwhile!

plan when expected GAT reduction > planning time
plan when it appears worthwhile!

plan when expected GAT reduction > planning time

GAT reduction depends on whether \( \hat{f}(\alpha) > \hat{f}(\beta) \) after search and if so, \( \hat{f}(\alpha) - \hat{f}(\beta) \)
plan when it appears worthwhile!

plan when expected GAT reduction > planning time

GAT reduction depends on whether \( \hat{f}(\alpha) > \hat{f}(\beta) \) after search and if so, \( \hat{f}(\alpha) - \hat{f}(\beta) \)

More precisely, if \( x_\alpha, x_\beta \) are possible \( \hat{f} \) values after search:

\[
b(x_\alpha, x_\beta) = \begin{cases} 
0 & \text{if } x_\alpha \leq x_\beta \\
 x_\alpha - x_\beta & \text{otherwise}
\end{cases}
\]

If \( P_{\hat{f}(n)} \) represents belief over future value,

\[
B = \int_{x_\alpha} P_{\hat{f}(\alpha)}(x_\alpha) \int_{x_\beta} P_{\hat{f}(\beta)}(x_\beta) b(x_\alpha, x_\beta) \, dx_\beta \, dx_\alpha
\]
how to estimate where $f(n)$ will be after search?
how to estimate where \( \hat{f}(n) \) will be after search?

if no search

belief about \( f^* \) = if search to goal

\( f(n) \quad \hat{f}(n) \)
how to estimate where $\hat{f}(n)$ will be after search?

- if no search
- intermediate search
- belief about $f^* = \text{if search to goal}$

$[\text{see paper for details}]$
variant of Dynamic $\hat{f}$ real-time search (Burns et al, SoCS-13)

1. until a goal is reached
2. best-first search on $\hat{f}$ until time bound
3. if identity is applicable and $B > t_{identity}$
4. $a \leftarrow$ identity
5. else
6. $a \leftarrow$ first action in best partial plan
7. update heuristic values
8. reset search
9. time bound $\leftarrow a$’s duration
10. start executing $a$
Results on Simple Benchmarks

Introduction

When to Plan More?
- Metareasoning
- Estimating Belief
- IMR
- Simple Problems
  - Large Problems 1
  - Large Problems 2

How Many Actions?

Conclusion

Sketches:

<table>
<thead>
<tr>
<th></th>
<th>A*</th>
<th>LSS-LRTA*</th>
<th>$\hat{f}$</th>
<th>IMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>cups</td>
<td>166</td>
<td>3,500</td>
<td>5,322</td>
<td>970</td>
</tr>
<tr>
<td>wall</td>
<td>102</td>
<td>523</td>
<td>717</td>
<td>101</td>
</tr>
<tr>
<td>slalom</td>
<td>177</td>
<td>382</td>
<td>638</td>
<td>161</td>
</tr>
<tr>
<td>uniform</td>
<td>29,578</td>
<td>3,195</td>
<td>2,997</td>
<td>2,997</td>
</tr>
</tbody>
</table>

IMR adapts from off-line to real-time!
Results on Larger Benchmarks (1/2)

Introduction

When to Plan More?
- Metareasoning
- Estimating Belief
- IMR
- Simple Problems
- Large Problems 1
- Large Problems 2

How Many Actions?

Conclusion

IMR approaches A*!

IMR adapts!
Introduction

When to Plan More?

- Metareasoning
- Estimating Belief
- IMR
- Simple Problems
- Large Problems 1
- Large Problems 2

How Many Actions?

Conclusion

Results on Larger Benchmarks (2/2)

Wheeler Ruml (UNH)

Concurrent Planning and Execution – 9 / 14
Consider each node along partial path
Stop at the first where planning is preferred
When combined with previous method: Mo’RTS
Results on Simple Benchmarks

Sketches:

<table>
<thead>
<tr>
<th>Problem</th>
<th>A*</th>
<th>LSS-LRTA*</th>
<th>$\hat{f}$</th>
<th>IMR</th>
<th>Mo’RTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cups</td>
<td>166</td>
<td>3,500</td>
<td>5,322</td>
<td>970</td>
<td>241</td>
</tr>
<tr>
<td>wall</td>
<td>102</td>
<td>523</td>
<td>717</td>
<td>101</td>
<td>140</td>
</tr>
<tr>
<td>slalom</td>
<td>177</td>
<td>382</td>
<td>638</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>uniform</td>
<td>29,578</td>
<td>3,195</td>
<td>2,997</td>
<td>2,997</td>
<td>2,997</td>
</tr>
</tbody>
</table>

Mo’RTS perhaps improves slightly over IMR
Results on Larger Benchmarks (1/2)

Introduction

When to Plan More?

How Many Actions?
- How Many?
- Simple Problems
- Large Problems 1
- Large Problems 2

Conclusion

Mo’RTS matches IMR

Mo’RTS matches IMR
Results on Larger Benchmarks (2/2)

Introduction

When to Plan More?

How Many Actions?
- How Many?
- Simple Problems
- Large Problems 1
- Large Problems 2

Conclusion

Possible slight improvement

Mo’RTS much better

Wheeler Ruml (UNH)
Conclusions

Objective: Minimize time to goal achievement

1. plan then act: Bugsy (Burns, Ruml, and Do, JAIR 2013)
2. concurrent planning and acting: Mo’RTS (this work)

Approach: Metareasoning

1. beautiful principle
2. provides state-of-the-art results in practice
3. should be integrated into the planner

Possible extensions

1. non-deterministic and partially-known settings
2. ‘not-quite-identity’ actions
3. plan-space planning

Practical metareasoning for adaptive deliberation!
Introduction

When to Plan More?

How Many Actions?

Conclusion

Extra Slides
- Assumptions
- DTA*
Assumptions

- inadmissible $\hat{h}$
- Gaussian belief
- linear variance reduction with lookahead
- estimate of future expansion delay
- cost of committing before frontier
- identity and length of commitment are separate decisions
- only consider acting at action end times
same basic principle
based on older RTA* instead of Dynamic $\hat{f}$
  - assumes disjoint subtrees beneath current actions
  - assumes admissible $h$
  - non-A* lookahead
estimates effect of search using training data