Distributional Methods for Heuristic Search

Tianyi Gu

Advisor: Wheeler RumI

University of New Hampshire
6th year CS PhD at UNH

Research: heuristic search and planning
  real-time heuristic search
  suboptimal search
  metareasoning

heuristic search can benefit from representing uncertainty

scalar heuristic $\rightarrow$ belief distribution that represents uncertainty
About Me

6th year CS PhD at UNH

Research: heuristic search and planning

real-time heuristic search
suboptimal search
metareasoning

heuristic search can benefit from representing uncertainty

scalar heuristic $\rightarrow$ belief distribution that represents uncertainty
What is Real-time Heuristic Search?

An example: path finding

agent performs search for a bounded time
What is Real-time Heuristic Search?

An example: path finding

agent performs search for a bounded time

agent

search frontier

goal
What is Real-time Heuristic Search?

An example: path finding

agent

best top level action

search frontier

agent commits to best action and executes

best node on the frontier

goal
What is Real-time Heuristic Search?

An example: path finding

agent commits to best action and executes
An example: path finding

agent commits to best action and executes

goal
What is Real-time Heuristic Search?

An example: path finding

- **Agent**
- **Goal**
- **Search frontier**
- **Best top level action**

**Online planning**: interleaving search and action execution

"Receding horizon control"
Real-time heuristic search:
return the next action within a time bound

Applications:
interacting with humans
dynamic environment
- autonomous vehicle
  inaccurate sensor
update model online
Real-time Search as Decision-making Under Uncertainty: The Nancy Framework
three phases:

1. Lookahead Phase:
   expands nodes with minimum $f$
   to explore the search space
three phases:

1. **Lookahead Phase:**
   - expands nodes with minimum $f$
   - to explore the search space

2. **Decision-making Phase:**
   - backup the minimum $f$ from search frontier (‘minimin’)
   - select top level action with minimum $f$ to execute
three phases:

1. Lookahead Phase:
   expands nodes with minimum $f$
to explore the search space
2. Decision-making Phase:
   backup the minimum $f$ from search frontier ('minimin')
   select top level action with minimum $f$ to execute
3. Learning Phase:
   update heuristic values
   (to escape local minima and avoid infinite loops)
A Classic Approach: LSS-LRTA* (Koenig&Sun 2008)

three phases:

1. Lookahead Phase:
   - expands nodes with minimum $f$
   - to explore the search space

2. Decision-making Phase:
   - backup the minimum $f$ from search frontier (‘minimin’)
   - select top level action with minimum $f$ to execute

3. Learning Phase:
   - update heuristic values
     - (to escape local minima and avoid infinite loops)

repeat until at a goal
three phases:

1. **Lookahead Phase:**
   expands nodes with minimum $f$
to explore the search space

2. **Decision-making Phase:**
   backup the minimum $f$ from search frontier (‘minimin’)
   select top level action with minimum $f$ to execute

3. **Learning Phase:**
   update heuristic values
   (to escape local minima and avoid infinite loops)

repeat until at a goal

derived from offline search, but optimal for online?
Should an agent at $A$ move to $B_1$ or $B_2$?
($x_i$ are unknown but i.i.d. uniform 0-1)

random tree domain (Pemberton & Korf 1995)

$f = g + h = g + 0$ is lower bound on optimal plan cost
Should an agent at $A$ move to $B_1$ or $B_2$?

($x_i$ are unknown but i.i.d. uniform 0-1)

decision theory says minimize expected value
lower bound: not suitable for rational action selection
Should an agent at $A$ move to $B_1$ or $B_2$? 
($x_i$ are unknown but i.i.d. uniform 0-1)
Should an agent at $A$ move to $B_1$ or $B_2$?

$(x_i$ are unknown but i.i.d. uniform 0-1)

$\hat{f}$ is expected plan cost

$f$ is not the answer: should minimize expected value!

plan under time pressure $\rightarrow$ bounded rationality
Lookahead Phase: A Troublesome Example

\begin{equation}
\hat{f}(\alpha) \quad \hat{f}(\beta)
\end{equation}

\(\hat{f}\) is expected value

Should an agent expand nodes under \(\alpha\) or \(\beta\)?
Lookahead Phase: A Troublesome Example

Introduction

Nancy

- Decision-making
- Lookahead
- The Beliefs
- Results

Conclusions

\[ \hat{f}('alpha') \hat{f}(\beta) \]

belief about \( \alpha \)

belief about \( \beta \)

\( \hat{f} \) is expected value

Should an agent expand nodes under \( \alpha \) or \( \beta \)?

\( \hat{f} \) is not the answer: what to do?

want to maximize value of information

need to consider uncertainty of estimates
Risk-based lookahead\(^1\):

- want to maximize value of information
- expand nodes which minimize expected regret
- relies on belief of values
- choose expansions that decrease uncertainty in beliefs

---

Backup Rules: Nancy

Nancy:
parent ← belief with minimum \( \hat{f} \) among successors
conveys an entire belief distribution
Heuristic values: scalar → probability distribution (belief)

But where do beliefs come from?
How to Form The Belief Distribution?

Heuristic values: scalar \( \rightarrow \) probability distribution (belief)

But where do beliefs come from?

Nancy:

truncated Gaussian based on \( \hat{f} \) and \( f \),

few parameters allows online learning

Data-Driven Nancy\(^2\):

expressive histogram,

many parameters requires offline learning

### Mean Solution Cost on Planning Domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Lookahead</th>
<th>LSS-LRTA*</th>
<th>Nancy</th>
<th>Nancy (DD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocksw.</td>
<td>100</td>
<td>46</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>36</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>30</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>Transport</td>
<td>100</td>
<td>631</td>
<td>615</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>519</td>
<td>559</td>
<td>485</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>499</td>
<td>567</td>
<td>422</td>
</tr>
<tr>
<td>Transport</td>
<td>100</td>
<td>48</td>
<td>40</td>
<td>31</td>
</tr>
<tr>
<td>(unit-cost)</td>
<td>300</td>
<td>47</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>35</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>Elevators</td>
<td>100</td>
<td>50</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>(unit-cost)</td>
<td>300</td>
<td>32</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>34</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

Both version of Nancy outperform conventional approach!
Conclusions
Examples of using distribution to guide search:
- real-time planning: Nancy, DDNancy
- suboptimal search: XES (IJCAI-21 paper 994)
- robotics: BEAST (IROS-17)
Examples of using distribution to guide search:
- real-time planning: Nancy, DDNancy
- suboptimal search: XES (IJCAI-21 paper 994)
- robotics: BEAST (IROS-17)

Exciting time in AI!
- Planning, RL, ML, Robotics
Examples of using distribution to guide search:
- real-time planning: Nancy, DDNancy
- suboptimal search: XES (IJCAI-21 paper 994)
- robotics: BEAST (IROS-17)

Exciting time in AI!
- Planning, RL, ML, Robotics

Much work needs to be done!
- data-driven + planning
- statistics + model-based approach
Questions?
Back-up Slides

- Completeness
- Search
expand under $\alpha$ or $\beta$?
expand under $\alpha$ or $\beta$?

need 2 things:
1) current beliefs
2) estimate of how beliefs might change with search
expand under $\alpha$ or $\beta$?

need 2 things:
1) current beliefs
2) estimate of how beliefs might change with search
Risk-based Lookahead Example

expand under $\alpha$ or $\beta$?

$$\textbf{Risk: } \text{expected regret if a suboptimal action is selected}$$

$\alpha$ is TLA with lowest expected value, other is $\beta$

$$\mathbb{E} \left[ f^*(\alpha) - f^*(\beta) \right] \mid \begin{cases} f^*(\beta) < f^*(\alpha) \text{ in cases when } \alpha \text{ not best} \\ \text{what is our regret} \end{cases}$$
Risk-based Lookahead Example

**Risk:** expected regret if a suboptimal action is selected

- $\alpha$ is TLA with lowest expected value, other is $\beta$

\[
\mathbb{E} \left[ f^*(\alpha) - f^*(\beta) \mid f^*(\beta) < f^*(\alpha) \right]
\]

- what is our regret
- in cases when $\alpha$ not best

Expand under $\alpha$ or $\beta$?
expand under $\alpha$ or $\beta$?

Risk-based Lookahead Example

$B_1$ $\alpha$ $A$ $\beta$ $B_2$

risk$_\alpha \approx \approx \approx$

risk$_\beta \approx \approx \approx$

expand under the TLA that minimizes risk!
expand under $\beta$!
**Lemma 1** Under assumptions of goal-awareness and finite state space, if a real-time search algorithm is incomplete, it must have a circulating set $S_\circ$.

**Lemma 5** Under our assumptions, a reasonable real-time search algorithm cannot have a circulating set.

**Theorem 1** Under our assumptions, a reasonable real-time search algorithm will eventually reach a goal.

**Lemma 7** Nancy is a reasonable real-time search algorithm.

**Lemma 8** LSS-LRTA* is a reasonable real-time search algorithm.

This proof applies to any LSS-LRTA*-style algorithm: no longer need heuristic consistency!
sliding tile puzzle
  uniform, heavy, inverse
pancake puzzle
  different size
racetrack
  reminiscent of autonomous driving
Comparison to IE and MCTS on Classic Search Domains

Introduction

Nancy

Conclusions

Back-up Slides
- Completeness
- Search

pancake puzzle

Start

Goal

Allowed Move
Nancy outperforms conventional approaches and MCTS\(^3\)

---

\(^3\) Real-time Planning as Data-driven Decision-making, Maximilian Fickert, Tianyi Gu, Leonhard Staut, Sai Lekyang, Wheeler Ruml, Joerg Hoffmann, and Marek Petrik, Bridging the Gap Between AI Planning and Reinforcement Learning (PRL), 2020.