Extreme Value Monte Carlo Tree Search for Classical Planning

Masataro Asai[1] and Stephen Wissow[2]

[1]MIT-IBM Watson AI Lab

[2]University of New Hampshire

Extreme Value Monte Carlo Tree Search for Classical Planning

Masataro Asai[1] and Stephen Wissow[2]

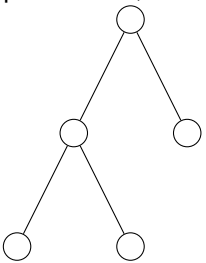
[1]MIT-IBM Watson AI Lab

[2]University of New Hampshire

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

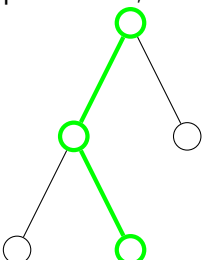
partial tree, mid-search:



Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:



1. **select** a leaf node by

$$\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2\log T}{t_i}}$$

$\hat{\mu}_i$: mean of child i

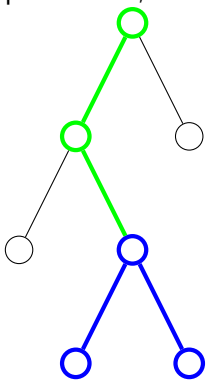
t_i : visit count of i

T : parent visit count

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

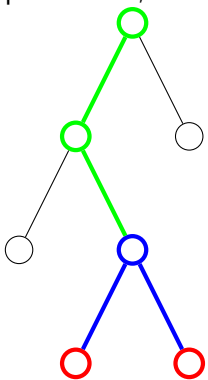


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2 \log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node

Monte Carlo Tree Search (MCTS/UCT)

Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

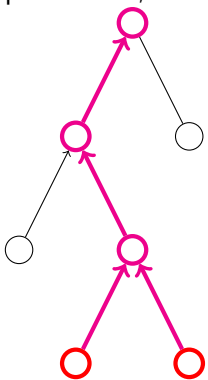


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2 \log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node
3. **evaluate** heuristics of children

Monte Carlo Tree Search (MCTS/UCT)

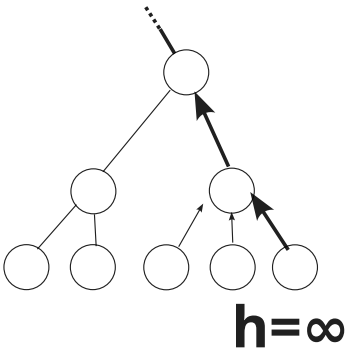
Auer, Cesa-Bianchi, and Fischer (2002, UCB1), Kocsis and Szepesvári (2006, UCT), Schulte and Keller (2014, THTS)

partial tree, mid-search:

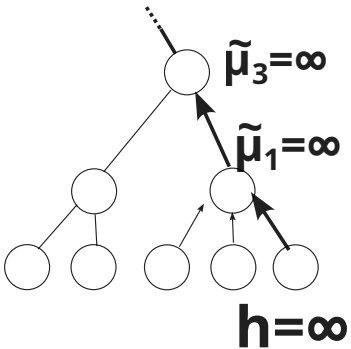


1. **select** a leaf node by
   $\text{LCB1}_i = \hat{\mu}_i - c\sqrt{\frac{2 \log T}{t_i}}$
   $\hat{\mu}_i$: mean of child i
   t_i : visit count of i
   T : parent visit count
2. **expand** a leaf node
3. **evaluate** heuristics of children
4. **backup** information to ancestors
   $t = \sum_i t_i$: Sum of children
   $\hat{\mu} = \frac{\sum_i t_i \hat{\mu}_i}{\sum_i t_i}$: Weighted avg

Average is Weird in Planning: What happens at a dead-end?

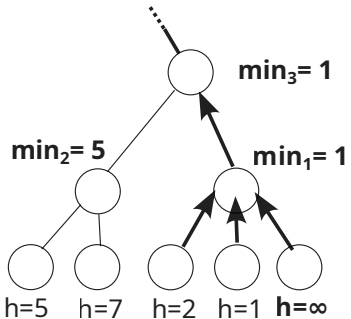


Average is Weird in Planning: What happens at a dead-end?



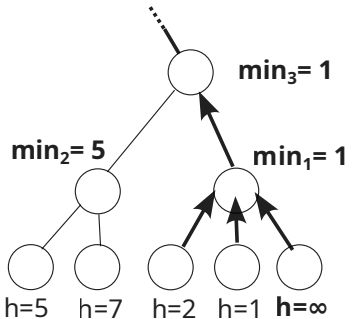
- **single** $h = \infty \rightarrow$ **all** $\hat{\mu} = \infty$

Average is Weird in Planning: What happens at a dead-end?



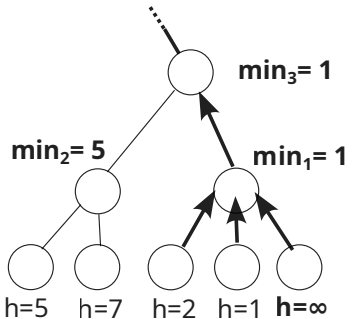
- **single** $h = \infty \rightarrow$ **all** $\hat{\mu} = \infty$
- $\min$ has no such problem

Average is Weird in Planning: What happens at a dead-end?



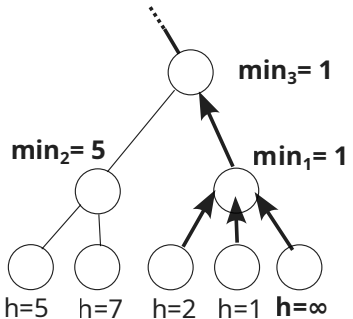
- **single** $h = \infty \rightarrow$ **all** $\hat{\mu} = \infty$
- $\min$ has no such problem
- GBFS: backup min h, select min h
  GUCT: backup avg h, select UCB1
  GUCT*: backup min h, select UCB1
  (Schulte and Keller 2014, THTS)
  However...

Average is Weird in Planning: What happens at a dead-end?



- **single** $h = \infty \rightarrow$ **all** $\hat{\mu} = \infty$
- $\min$ has no such problem
- GBFS: backup min h, select min h
  GUCT: backup avg h, select UCB1
  GUCT*: backup min h, select UCB1
  (Schulte and Keller 2014, THTS)
  However...
- $\min$ **lacks statistical interpretation**

Average is Weird in Planning: What happens at a dead-end?



- **single $h = \infty \rightarrow$ all $\hat{\mu} = \infty$**
- $\min$ has no such problem
- GBFS: backup min h, select min h
  GUCT: backup avg h, select UCB1
  GUCT*: backup min h, select UCB1
  (Schulte and Keller 2014, THTS)
  However...
- $\min$ **lacks statistical interpretation**
- GUCT/* remove ∞ from tree
  ... **also lacks statistical interpretation**

The average is weird, but how to use the minimum with statistical rigor?

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average
is the **Central Limit Theorem (CLT).**

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average
is the **Central Limit Theorem (CLT).**

The statistical theory of the minimum (or maximum)?

The average is weird, but how to use the minimum with statistical rigor?

The statistical theory of the average
is the **Central Limit Theorem (CLT).**

The statistical theory of the minimum (or maximum)?
It's **Extreme Value Theory (EVT)!**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. Maximum water level in rivers

Extreme Value Theory (EVT)

Safety-critical applications: e.g. Maximum water level in rivers

- average water level $\rightarrow$ Gaussian distribution

Extreme Value Theory (EVT)

Safety-critical applications: e.g. Maximum water level in rivers

- average water level $\rightarrow$ Gaussian distribution
- **Exceedance over safety limit $\rightarrow$ Generalized Pareto (GP) distribution**

Extreme Value Theory (EVT)

Safety-critical applications: e.g. Maximum water level in rivers

- average water level $\rightarrow$ Gaussian distribution
- **Exceedance over safety limit $\rightarrow$ Generalized Pareto (GP) distribution**

$$\mathrm{GP}(\mathrm{x} \mid \theta, \sigma, \xi) = \langle \text{complicated math} \rangle \ (\mathrm{x} > \theta)^{*} \textbf{ for threshold } \theta$$

Extreme Value Theory (EVT)

Safety-critical applications: e.g. Maximum water level in rivers

- average water level $\rightarrow$ Gaussian distribution
- **Exceedance over safety limit $\rightarrow$ Generalized Pareto (GP) distribution**

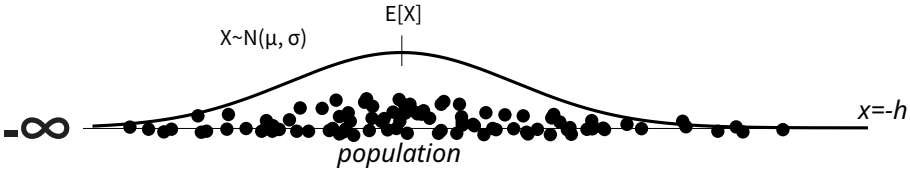
$$\mathrm{GP}(\mathrm{x} \mid \theta, \sigma, \xi) = \langle \text{complicated math} \rangle \ (\mathrm{x} > \theta)^{*} \textbf{ for threshold } \theta$$

***justifies removing dead-ends:**

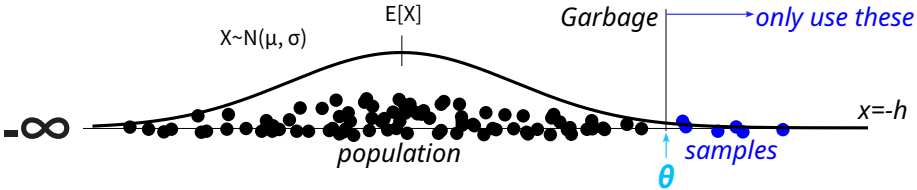
- define $x = -h$
- dead-end: $h = \infty \Rightarrow x < \theta$

Backup = Fitting a distribution

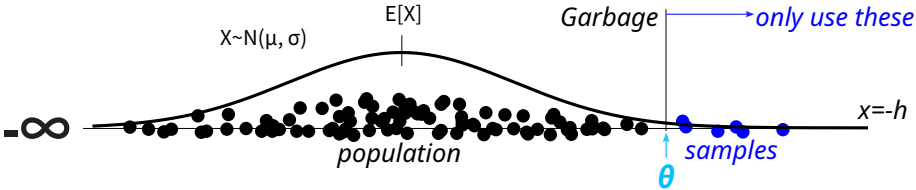
Backup = Fitting a distribution



Backup = Fitting a distribution

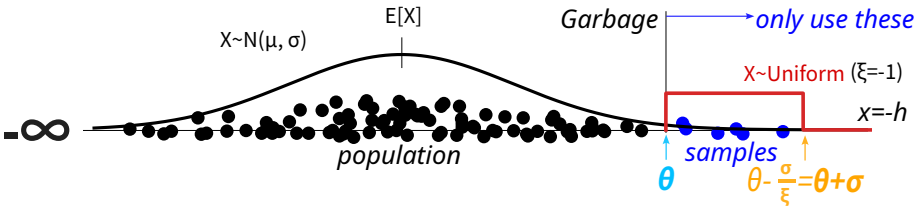


Backup = Fitting a distribution



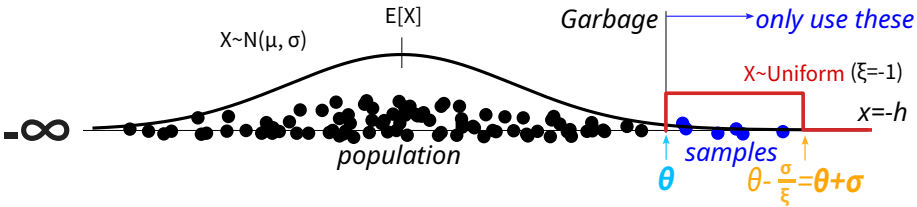
- We fit **exceedance over** $\boldsymbol{\theta} = -\max_{\boldsymbol{s}} h(\boldsymbol{s})$ to $\mathrm{GP}(\boldsymbol{\theta}, \sigma, \xi)$

Backup = Fitting a distribution



- We fit **exceedance over** $\boldsymbol{\theta} = -\max_{\boldsymbol{s}} h(\boldsymbol{s})$ to $\mathrm{GP}(\boldsymbol{\theta}, \sigma, \xi)$
- We use a **special case** $\mathrm{GP}(\boldsymbol{\theta}, \sigma, -1)$
$= \mathrm{Uniform}(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\sigma}) : \boldsymbol{\sigma} = \max_{\boldsymbol{s}} h(\boldsymbol{s}) - \min_{\boldsymbol{s}} h(\boldsymbol{s})$

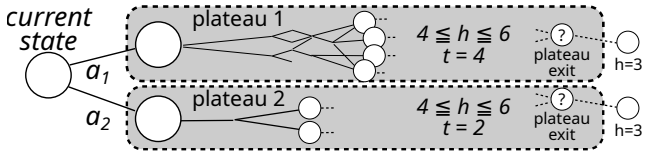
Backup = Fitting a distribution



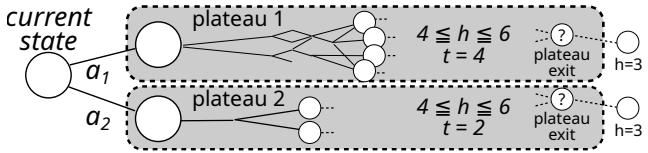
- We fit **exceedance over** $\boldsymbol{\theta} = -\max_{\boldsymbol{s}} h(\boldsymbol{s})$ to $\mathrm{GP}(\boldsymbol{\theta}, \sigma, \xi)$
- We use a **special case** $\mathrm{GP}(\boldsymbol{\theta}, \sigma, -1)$
  $= \mathrm{Uniform}(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\sigma}) : \boldsymbol{\sigma} = \max_{\boldsymbol{s}} h(\boldsymbol{s}) - \min_{\boldsymbol{s}} h(\boldsymbol{s})$
- We define a **new Bandit**:

$$\text{LCB1-Uniform}_i = \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6 t_i \log T}$$
$$\hat{u}_i = \max_{\boldsymbol{s}} h(\boldsymbol{s}), \qquad \hat{l}_i = \min_{\boldsymbol{s}} h(\boldsymbol{s})$$

LCB1-Uniform: Spread-Aware to Handle Plateaus

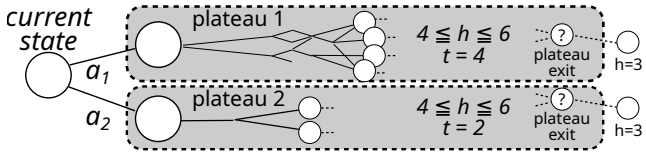


LCB1-Uniform: Spread-Aware to Handle Plateaus



$$\text{LCB1-Uniform}_i = \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6t_i \log T}$$

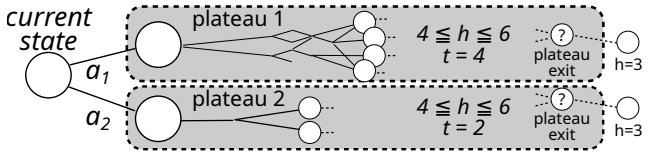
LCB1-Uniform: Spread-Aware to Handle Plateaus



$$\text{LCB1-Uniform}_i = \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6t_i \log T}$$

- penalize narrow spread: **avoid plateaus**
  (more generally, regions with small h variations)

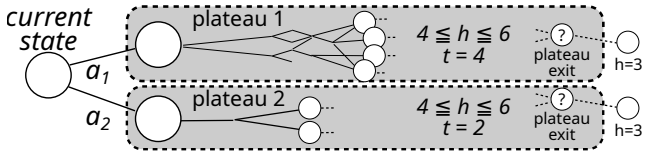
LCB1-Uniform: Spread-Aware to Handle Plateaus



$$\text{LCB1-Uniform}_i = \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6t_i \log T}$$

- penalize narrow spread: **avoid plateaus**
  (more generally, regions with small h variations)
- prefer wider spread: **smaller h more likely in the future**

LCB1-Uniform: Spread-Aware to Handle Plateaus



$$\text{LCB1-Uniform}_i = \frac{\hat{u}_i + \hat{l}_i}{2} - (\hat{u}_i - \hat{l}_i)\sqrt{6t_i \log T}$$

- penalize narrow spread: **avoid plateaus**
  (more generally, regions with small h variations)
- prefer wider spread: **smaller h more likely in the future**
- one plateau at a time: **exit plateaus quickly** ~~explore other plateaus first~~

Best Expansion Order ($\leq$10k node evals, 24 IPC domains)

$h =$	h^{FF}	h^{add}	$h^{\max}$	h^{GC}	h^{FF}+PO	h^{FF}+DE+PO
GBFS (Pyperplan/FD)	538/539	518/517	224/226	354/349	†/539	†/‡
Softmin-Type(h)	576	542.6	297.2	357.6	575.8	‡
GUCT	412	397.8	228.4	285.2	454.2	440.4
GUCT*	459.4	480.8	242.2	312.2	496.2	471.8
GUCT-Normal	283.4	265	212	233.4	372.4	381.6
GUCT*-Normal	318.8	300	215.2	246.2	378.05	386.9
GUCT-Normal2	582.95	538	316.6	380.6	623.2	581.8
GUCT*-Normal2	567.2	533.8	263	341.2	619.8	570.6
GUCT-Uniform (ours)	**606.4**	**563.4**	**455.6**	**492.2**	**635.6**	**600.8**
CHK-Uniform	375.4	338.8	224.8	296.6	454.8	458.2
GUCT+-Normal2	578	550.4	442.4	490.6	630.6	582.2
MaxSearch	253.75	243.4	260	255.2	368.6	355.6
RobustUCT	267.8	270.8	234	231.8	403	435.2
ThresholdAscent	162.4	163.8	170.4	164.4	165.8	172.2

Fast Runtimes: **Best Agile IPC Score** (h^{FF}, IPC-18, Fast Downward)

Instances solved

domain	GBFS	SM	N2	**Uni**
agricola	9.0	10.2	9.4	**11.6**
caldera	4.0	**7.0**	6.4	5.8
data-net	4.0	**8.4**	8.2	7.0
flashfill	**9.0**	8.8	7.2	6.8
nurikabe	7.0	6.2	**8.4**	7.6
org-syn	9.0	8.8	**9.2**	6.0
settlers		**5.4**	2.4	2.6
snake	5.0	5.0	15.4	**19.0**
spider	8.0	8.2	**9.2**	8.6
termes	**12.0**	11.6	5.8	5.0
total	67.0	79.6	**81.6**	80.0

IPC score

domain	GBFS	SM	N2	**Uni**
agricola	1.9	3.4	2.0	**6.1**
caldera	2.9	5.0	5.1	**5.3**
data-net	3.5	4.5	**5.6**	4.9
flashfill	6.2	**6.8**	5.4	4.8
nurikabe	6.4	5.4	**6.9**	6.5
org-syn	**7.2**	6.8	6.7	5.0
settlers		**2.6**	1.6	2.3
snake	2.9	3.3	9.1	**12.5**
spider	2.2	3.1	**3.4**	**3.4**
termes	**6.4**	**6.2**	2.6	2.5
total	39.5	47.0	48.6	**53.2**

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**
2. **The statistical theory of maximum: Extreme Value Theory**

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**
2. **The statistical theory of maximum: Extreme Value Theory**
3. **Solution**: EVT $\rightarrow$ Generalized Pareto $\rightarrow$ Uniform $\rightarrow$ Uniform bandit

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**
2. **The statistical theory of maximum: Extreme Value Theory**
3. **Solution**: EVT $\rightarrow$ Generalized Pareto $\rightarrow$ Uniform $\rightarrow$ Uniform bandit
4. **Result**: Good! Fewer expansions, faster runtimes

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**
2. **The statistical theory of maximum: Extreme Value Theory**
3. **Solution**: EVT $\rightarrow$ Generalized Pareto $\rightarrow$ Uniform $\rightarrow$ Uniform bandit
4. **Result**: Good! Fewer expansions, faster runtimes

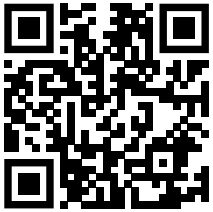
Find our full paper:

https://arxiv.org/abs/2405.18248

Conclusion

1. **Conundrum**: average is weird, min is good, want to prune $h = \infty$. **How?**
2. **The statistical theory of maximum: Extreme Value Theory**
3. **Solution**: EVT $\rightarrow$ Generalized Pareto $\rightarrow$ Uniform $\rightarrow$ Uniform bandit
4. **Result**: Good! Fewer expansions, faster runtimes

Find our full paper:
https://arxiv.org/abs/2405.18248



ECAI-24 Outstanding Paper:
https://cs.unh.edu/~sjw1000