Planner Metrics Should Satisfy Independence of Irrelevant Alternatives

Jendrik Seipp
July 12, 2019

University of Basel, Switzerland
Independence of irrelevant alternatives (IIA)

- one of four criteria from Arrow’s impossibility theorem
- decision whether $A > B$ or $A < B$ is irrelevant from $C$
Independence of irrelevant alternatives (IIA)

- one of four criteria from Arrow’s impossibility theorem
- decision whether A > B or A < B is irrelevant from C
- important for planner metrics, but some violate it
IPC satisficing track

\[
sat(P, \pi) = \begin{cases} 
\frac{Cost^*(\pi)}{Cost(P, \pi)} & \text{if solved} \\
0 & \text{if unsolved}
\end{cases}
\]

- total score: sum of task scores
- \(Cost^*(\pi)\) is the cost of a reference plan
IPC satisficing track

\[
sat(P, \pi) = \begin{cases} 
\frac{\text{Cost}^*(\pi)}{\text{Cost}(P, \pi)} & \text{if solved} \\
0 & \text{if unsolved}
\end{cases}
\]

- total score: sum of task scores
- \(\text{Cost}^*(\pi)\) is the cost of a reference plan
- if reference plans are optimal, sat satisfies IIA
IPC satisficing track

\[
\text{sat}(P, \pi) = \begin{cases} 
\frac{\text{Cost}^*(\pi)}{\text{Cost}(P, \pi)} & \text{if solved} \\
0 & \text{if unsolved}
\end{cases}
\]

- total score: sum of task scores
- \(\text{Cost}^*(\pi)\) is the cost of a reference plan
- if reference plans are optimal, sat satisfies IIA
- if reference plans can come from competitors, sat does not satisfy IIA
## IPC satisficing track – example

<table>
<thead>
<tr>
<th>Cost</th>
<th>R</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

sat

<table>
<thead>
<tr>
<th>sat</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2/5</td>
<td>2/4</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>4/4</td>
<td>4/5</td>
</tr>
</tbody>
</table>

$\sum$

| $\sum$ | 1.4 | 1.3 |

$\rightarrow$ A > B
### IPC satisficing track – example

<table>
<thead>
<tr>
<th>Cost</th>
<th>R</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sat</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2/5</td>
<td>2/4</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>4/4</td>
<td>4/5</td>
</tr>
</tbody>
</table>

\[ \sum \quad 1.4 \quad 1.3 \]

$\rightarrow A > B$

### Example

<table>
<thead>
<tr>
<th>sat</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2/5</td>
<td>2/4</td>
<td>2/5</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>1/4</td>
<td>1/5</td>
<td>1/1</td>
</tr>
</tbody>
</table>

\[ \sum \quad 0.65 \quad 0.7 \quad 1.4 \]

$\rightarrow B > A$

Use optimal planners or domain-specific solvers to find good reference plans.
### IPC satisficing track – example

<table>
<thead>
<tr>
<th>Cost</th>
<th>R</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi_1)</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(\pi_2)</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sat</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi_1)</td>
<td>2/5</td>
<td>2/4</td>
</tr>
<tr>
<td>(\pi_2)</td>
<td>4/4</td>
<td>4/5</td>
</tr>
</tbody>
</table>

\[\sum\] 1.4 1.3

\(\rightarrow A > B\)

<table>
<thead>
<tr>
<th>sat</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi_1)</td>
<td>2/5</td>
<td>2/4</td>
<td>2/5</td>
</tr>
<tr>
<td>(\pi_2)</td>
<td>1/4</td>
<td>1/5</td>
<td>1/1</td>
</tr>
</tbody>
</table>

\[\sum\] 0.65 0.7 1.4

\(\rightarrow B > A\)

\(\rightarrow\) use **optimal planners** or **domain-specific solvers** to find good reference plans
T*(\(\pi\)): minimum runtime of all participating planners

\[
\text{agl}_{2014}(P, \pi) = \begin{cases} 
1/(1 + \log_{10} \frac{T(P, \pi)}{T^{*}(\pi)}) & \text{if } T(P, \pi) \leq 300 \\
0 & \text{otherwise}
\end{cases}
\]
$T^*(\pi)$: minimum runtime of all participating planners

$$\text{agl}_{2014}(P, \pi) = \begin{cases} 
\frac{1}{1 + \log_{10} \frac{T(P, \pi)}{T^*(\pi)}} & \text{if } T(P, \pi) \leq 300 \\
0 & \text{otherwise}
\end{cases}$$
IPC agile track

$T^*(\pi)$: minimum runtime of all participating planners

$$\text{agl}_{2014}(P, \pi) = \begin{cases} 
1/(1 + \log_{10} \frac{T(P, \pi)}{T^*(\pi)}) & \text{if } T(P, \pi) \leq 300 \\
0 & \text{otherwise}
\end{cases}$$

$$\text{agl}_{2018}(P, \pi) = \begin{cases} 
1 & \text{if } T(P, \pi) < 1 \\
1 - \frac{\log(T(P, \pi))}{\log(300)} & \text{if } 1 \leq T(P, \pi) \leq 300 \\
0 & \text{if } T(P, \pi) > 300
\end{cases}$$

→ use $\text{agl}_{2018}$ in future agile tracks
Sparkle planning challenge

• new planning competition in 2019
• “analyse the contribution of each planner to the real state of the art”
• measure marginal contribution of each planner $P$ to a portfolio selector over planners $S$

$$\text{sparkle}(P, \pi) = \begin{cases} 
\log_{10} \frac{\text{par10}(S \setminus \{P\})}{\text{par10}(S)} & \text{if } \text{par10}(S \setminus \{P\}) > \text{par10}(S) \\
0 & \text{otherwise}
\end{cases}$$
Sparkle planning challenge

- new planning competition in 2019
- “analyse the contribution of each planner to the real state of the art”
- measure marginal contribution of each planner $P$ to a portfolio selector over planners $S$

$$\text{sparkle}(P, \pi) = \begin{cases} 
\log_{10} \frac{\text{par}_{10}(S \setminus \{P\})}{\text{par}_{10}(S)} & \text{if } \text{par}_{10}(S \setminus \{P\}) > \text{par}_{10}(S) \\
0 & \text{otherwise}
\end{cases}$$
Sparkle planning challenge

- new planning competition in 2019
- “analyse the contribution of each planner to the real state of the art”
- measure marginal contribution of each planner $P$ to a portfolio selector over planners $S$

\[
\text{sparkle}(P, \pi) = \begin{cases} 
\log_{10} \frac{\text{par}10(S \setminus \{P\})}{\text{par}10(S)} & \text{if } \text{par}10(S \setminus \{P\}) > \text{par}10(S) \\
0 & \text{otherwise} 
\end{cases}
\]

- focuses on coverage
- uses runtime to break ties
Sparkle planning challenge

- new planning competition in 2019
- “analyse the contribution of each planner to the real state of the art”
- measure marginal contribution of each planner $P$ to a portfolio selector over planners $S$

$$\text{sparkle}(P, \pi) = \begin{cases} 
\log_{10} \frac{\text{par10}(S \{P\})}{\text{par10}(S)} & \text{if } \text{par10}(S \{P\}) > \text{par10}(S) \\
0 & \text{otherwise}
\end{cases}$$

- focuses on coverage
- uses runtime to break ties
- removing which planner decreases coverage the most?
• 100 tasks
• planner A solves 1 task $\pi$
• planners B and C solve 99 tasks but fail to solve $\pi$
Sparkle planning challenge – example

• 100 tasks
• planner A solves 1 task $\pi$
• planners B and C solve 99 tasks but fail to solve $\pi$
• $\{A, B\} \rightarrow B > A$
• $\{A, B, C\} \rightarrow A > B$
Sparkle planning challenge – problems of the metric

• penalizes similar planners
• easily gameable: submit several “dummy” planners and one “real” planner (leader board, IPC planners available)
• penalizes collaboration, favors closed-source planners
• discourages submitting multiple planners
Sparkle planning challenge – suggestion

• IIA: use fixed portfolio of baseline planners
Summary

- IIA is critical for evaluation metrics
- several planner metrics do not satisfy IIA
- there are alternatives that do satisfy IIA