TCP-nets for Preferences over Sets

R. Brafman, C. Domshlak, S. Shimony, & Y. Silver
Standard setting

Using a given attribution $X = \{X_1, \ldots, X_n\}$ of the alternatives $O = \times \text{Dom}(X_i)$, the user expresses her preference over $O$. 

Preference Query

Query result

Query evaluation

DBMS

CSP
Preferences over Sets

Using a given attribution $X = \{X_1, \ldots, X_n\}$ of the alternatives $O = \times \text{Dom}(X_i)$, the user expresses her preference over the powerset $2^O$. 

Preference Query

Query result

Query evaluation

DBMS

CSP
Preferences over Sets

The optimal subset selection problem:

• Content of a news page.
• IJCAI program committee, football team, etc.
• Commercial deals.
• ...

Is there a problem?

• Stay in the standard setting and output the top-most preferred alternatives from the database.
  • Typically inappropriate.

• Model each alternative in the database as a binary (in/out) variable.
  • Too many variables! (preference elicitation ...)
  • Very complicated preferential dependencies.
  • What if the database is not static?
Our desiderata

Static preference specification approach that can be used for dynamically selecting a subset of objects, where the whole set of available objects is unknown at the time of preference specification.

For example, while trying to configure the optimal football team, we would like to be able to characterize the preferred properties of a team, and then to select from a given set of individuals the best subset for our task.
General framework

1. Obtain from the user properties of sets of objects that affect her preferences over these sets.

2. Allow the user to express preference statements in terms of these properties.

3. Construct a preference representation model over these properties that captures the information provided by the statements in (2).
Set properties

Informally:

“... at least one object with \( C=c_1 \), and \( D=d_3 \) or \( D=d_5 \) ...”

“... the number of objects with \( C=c_1 \) should be ...”

- Basic propositions \( \overline{X} = \{ X = x \mid X \in X, x \in \text{Dom}(X) \} \),
- Propositional language \( \mathcal{L}_{\overline{X}} \).
- For \( o \in O, \phi \in \mathcal{L}_{\overline{X}} \), it makes sense to write \( o \models \phi \).

\[ \langle |\phi| \rangle \]
\[ \langle |\phi| \rangle(O) = |\{ o \in O | o \models \phi \}| \]

\[ \langle |\phi| \text{ REL } k \rangle \]
\[ \langle |\phi| \text{ REL } k \rangle(O) \Leftrightarrow |\{ o \in O | o \models \phi \}| \text{ REL } k \]
Example: Newsletter “optimization”

\[
\begin{align*}
P_1 &= \langle (\text{format} = \text{news}) \rangle \\
P_2 &= \langle (\text{emotion} = \text{neutral}) \lor (\text{emotion} = \text{negative}) \mid \leq 2 \rangle \\
P_3 &= \langle (\text{country} = \text{Iraq}) \land (\text{topic} = \text{politics}) \mid \geq 2 \rangle \\
P_4 &= \langle (\text{topic} = \text{culture}) \lor (\text{emotion} = \text{positive}) \mid \geq 1 \rangle
\end{align*}
\]

- **Format**: news, interview, opinion, etc.
- **Country**: Iraq, U.S.A, Italy, etc.
- **Topic**: politics, weather, economy, culture, etc.
- **Emotion**: positive, negative, neutral, etc.
Example: Newsletter “optimization”

\[ P_1 = \langle |(\text{format} = \text{news})| \rangle \]
\[ P_2 = \langle |(\text{emotion} = \text{neutral}) \lor (\text{emotion} = \text{negative})| \leq 2 \rangle \]
\[ P_3 = \langle |(\text{country} = \text{Iraq}) \land (\text{topic} = \text{politics})| \geq 2 \rangle \]
\[ P_4 = \langle |(\text{topic} = \text{culture} \lor (\text{emotion} = \text{positive})| \geq 1 \rangle \]

\[ O = \]

<table>
<thead>
<tr>
<th>format</th>
<th>country</th>
<th>topic</th>
<th>emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>Iraq</td>
<td>politics</td>
<td>neutral</td>
</tr>
<tr>
<td>news</td>
<td>U.S.A.</td>
<td>weather</td>
<td>negative</td>
</tr>
<tr>
<td>interview</td>
<td>Iraq</td>
<td>economy</td>
<td>positive</td>
</tr>
<tr>
<td>opinion</td>
<td>France</td>
<td>culture</td>
<td>positive</td>
</tr>
</tbody>
</table>

\[ P_1(O) = 2 \]
\[ P_2(O) = \text{true} \]
\[ P_3(O) = \text{false} \]
\[ P_4(O) = \text{true} \]
Problem reduction

\[ \mathcal{P} = \{P_1, \ldots, P_m\} \]

- Each property \( P_i \) can be seen as a (boolean/multi-valued) variable.
- Each subset \( O \subseteq \mathcal{O} \) provides a complete assignment to \( \mathcal{P} \).

- Abstractly, we can view each \( O \subseteq \mathcal{O} \) as a vector of values, \( p_O \), for \( P_1, \ldots, P_m \).
- Correspondence between \( 2^\mathcal{O} \) and the abstract set of outcomes \( \mathcal{O}_\mathcal{P} = \times_{\mathcal{P}} \mathcal{D}(P_i) \).
- Any preference order over \( \mathcal{O}_\mathcal{P} \) implicitly induces a preference order over \( 2^\mathcal{O} \).

By using \( \mathcal{O}_\mathcal{P} \) instead of \( 2^\mathcal{O} \) we have reduced our problem of specifying preferences over subsets to that of specifying preferences over attributed objects.
Background: TCP-nets

Day of the flight

<table>
<thead>
<tr>
<th>1d</th>
<th>m &gt; n</th>
</tr>
</thead>
<tbody>
<tr>
<td>2d</td>
<td>n &gt; m</td>
</tr>
</tbody>
</table>

Departure time

Stop-overs

<table>
<thead>
<tr>
<th>m</th>
<th>1s &gt; 0s</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>0s &gt; 1s</td>
</tr>
</tbody>
</table>

Class

Airline

<table>
<thead>
<tr>
<th>m ∧ ba</th>
<th>b &gt; e</th>
</tr>
</thead>
<tbody>
<tr>
<td>n ∧ klm</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m ∧ klm</th>
<th>S △ C</th>
</tr>
</thead>
<tbody>
<tr>
<td>n ∧ ba</td>
<td>S △ C</td>
</tr>
<tr>
<td>m ∧ ba</td>
<td>C △ S</td>
</tr>
</tbody>
</table>
1. Her preference on the amount of non-positive articles (P2) should be considered as less important than her preference on the amount of news in the article collection (P1),

2. Her preference on the amount of culture related articles (P4) should be considered as less important than her preference on the amount of articles on the political situation in Iraq (P3),

3. Her preference on having at least two articles on the political situation in Iraq (P3) depends on the amount of all news articles going into the newsletter (P1), and

4. Her preference on having some culture related articles (P4) depends on the amount of all news articles (P1) and on whether there are more than two non-positive articles in the issue or not (P2).
Example

1. Her preference on the amount of non-positive articles ($P_2$) should be considered as less important than her preference on the amount of news in the article collection ($P_1$),

2. Her preference on the amount of culture related articles ($P_4$) should be considered as less important than her preference on the amount of articles on the political situation in Iraq ($P_3$),

3. Her preference on having at least two articles on the political situation in Iraq ($P_3$) depends on the amount of all news articles going into the newsletter ($P_1$), and

4. Her preference on having some culture related articles ($P_4$) depends on the amount of all news articles ($P_1$) and on whether there are more than two non-positive articles in the issue or not ($P_2$).
Optimal Subset Selection

Input:

- A TCP-net \( N \) over \( \mathcal{P} = \{P_1, \ldots, P_m\} \).
- Currently available set of objects \( \mathcal{O} \).

Output:

- \( O \in \mathcal{O} \), such that for all other \( O' \in \mathcal{O} \) we have \( N \not\models p_{O'} \succ p_O \).

One of the attractive properties of TCP-nets is that it comes with an algorithm for constrained optimization that is guaranteed to examine solutions to hard constraints in an order consistent with the preference relation induced by the network.

[BD-uai02]
Algorithms

Two very different algorithms based on (and preserving the anytime property of) the standard algorithm for COP with TCP-nets.

• **Subset Selection by Hypothesis Refinement.**

• **Preference-based CSP Generation.**
Hypothesis Refinement

\[ P_1 = 0 \quad P_1 = 1 \]
Hypothesis Refinement

\[ P_1 = 0 \quad P_1 = 1 \]
\[ P_2 = 0 \quad P_2 = 1 \]
Hypothesis Refinement

$P_1 = 0$

$P_1 = 1$

$P_2 = 0$

$P_2 = 1$
Hypothesis Refinement

$P_1 = 0$ $P_1 = 1$

$P_2 = 0$ $P_2 = 1$

Tips:

*BDD*-style set representation.
Preference-based CSP Generation

\[
\text{CSP}(p, \mathcal{O}) \rightarrow \{\mathcal{O}\}_p
\]
Preference-based CSP Generation

Tips:
CSP solvers optimized for \textit{cardinality constraints}.
Summary and Future work

- Preference-based subset optimization.
  - Problem formalization.
  - Modeling and solving with TCP-nets.
  - Lots to be done!

- The problem is practically significant, and seems to pose numerous modeling and algorithmic challenges.
- Different domains - different challenges?
- Tell us about other real-life needs that map well to this subset optimization problem!