Improving Web Catalog Design for Easy Product Search

I. Robert Chiang
(with Manuel A. Nunez)

OPIM Department
University of Connecticut
Outline

- Improving web site navigation
- Brief literature
- Two scenarios
  - Optimize under fixed topology
  - Complete catalog redesign
- One-for-all vs. tailored solutions
- Remarks and future work
The focus here is not about page aesthetics but about topological issues. Especially applicable to the online catalog. The online catalog is one of the many headaches for “dot-goners”.

- Existing means of navigation
  - side bar
  - site search engine
  - hyperlinks

- Our goal is to automate the hyperlink design
Research Objectives

- Incorporate browsing logs and purchase histories to
  - Facilitate product search and bundling
  - Allow for customization and personalization
Prior Work

- Web page design
  - general quality (Aladwani and Palvia 2002)
  - usability and human factors (usability.org & humanfactors.com)
- Clicks stream analysis
  - Advertisement design (Karuga et al. 2001)
  - CRM (Padmanabhan and Tuzhilin 2003)
- Collaborative filtering
  - RACOFI (Anderson et al. 2003), KDD
- Site search engine design
Scenario 1: Fixed Topology

(The FT problem)
Solution Procedure

- Find out the correlations among *items*
  - $F$: item $i$ and $j$ bought or viewed in the same session

- Figure out the distance among *catalog pages*
  - $D$: number of clicks between page $k$ and $l$

- Place related items *close* to each other
  - $X$: assigning items to pages
Measuring the Distance Between Pages

- A 7-node catalog network
Measuring the Distance Between Pages (cont.)

- **Optimistic (shortest path)**

\[ D^B = \begin{bmatrix}
0 & 1 & 1 & 2 & 2 & 2 & 2 \\
1 & 0 & 2 & 1 & 1 & 3 & 3 \\
1 & 2 & 0 & 3 & 3 & 1 & 1 \\
2 & 1 & 3 & 0 & 2 & 4 & 4 \\
2 & 1 & 3 & 2 & 0 & 4 & 4 \\
2 & 3 & 1 & 4 & 4 & 0 & 2 \\
2 & 3 & 1 & 4 & 4 & 2 & 0
\end{bmatrix}. \]

- **Pessimistic (random walk)**

\[ D^W = \begin{bmatrix}
0 & 7 & 7 & 14 & 14 & 14 & 14 \\
5 & 0 & 12 & 9 & 9 & 19 & 19 \\
5 & 12 & 0 & 19 & 19 & 9 & 9 \\
6 & 1 & 13 & 0 & 10 & 20 & 20 \\
6 & 1 & 13 & 10 & 0 & 20 & 20 \\
6 & 13 & 1 & 20 & 20 & 0 & 10 \\
6 & 13 & 1 & 20 & 20 & 10 & 0
\end{bmatrix}. \]

\[ D := D(\theta) = \theta D^B + (1 - \theta) D^W. \]
Assigning Items to Pages

- Minimizing the average click count

\[
\text{(QAP) } \min z(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} F_{ij} \sum_{k=1}^{n} \sum_{l=1}^{n} D_{kl} X_{ik} X_{jl},
\]

s.t.

\[
\sum_{l=1}^{n} X_{il} = 1, \forall i,
\]

\[
\sum_{i=1}^{n} X_{il} = 1, \forall l,
\]

\[
X_{il} \in \{0, 1\}, \forall i, l.
\]
Solution Procedure

- For a 12-node problem
  - Enumeration
  - GA (20 chromosomes, 100 gen.)
  - IP heuristic (nearest neighbor)

- For a 120-node problem
  - GA (100 chromosomes, 100 gen.)
  - IP heuristic
The Twelve Items

- Lighting and household tools and supplies from HD

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Cluster</th>
<th>Item</th>
<th>Description</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Floodlight bulbs</td>
<td>A</td>
<td>7</td>
<td>Tape rolls</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>Bulb changer kit</td>
<td>B</td>
<td>8</td>
<td>Vinyl tape</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>Indoor timer</td>
<td>B</td>
<td>9</td>
<td>All purpose tool</td>
<td>D</td>
</tr>
<tr>
<td>4</td>
<td>Electrical tape</td>
<td>C</td>
<td>10</td>
<td>Fish tape</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>Wire connector kit</td>
<td>C</td>
<td>11</td>
<td>Fish tape leader</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>Cable tie kit</td>
<td>C</td>
<td>12</td>
<td>Scissors with stripping</td>
<td>D</td>
</tr>
</tbody>
</table>
Current Topology
Numerical Comparisons for 12-node Problem

- Objective values

<table>
<thead>
<tr>
<th></th>
<th>$\theta$</th>
<th>Objective value ($Z$)</th>
<th>Relative error**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enumeration</td>
<td>0.5</td>
<td>8.324*</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>2.775*</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>2.082*</td>
<td>N.A.</td>
</tr>
<tr>
<td>GA</td>
<td>0.5</td>
<td>8.324</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>2.775</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>2.082</td>
<td>0.00%</td>
</tr>
<tr>
<td>IP Heuristic</td>
<td>0.5</td>
<td>9.363</td>
<td>12.49%</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>3.246</td>
<td>16.97%</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>2.482</td>
<td>19.21%</td>
</tr>
</tbody>
</table>

- Computation time
  - Four hours for enumeration vs. a few seconds for GA and IP heuristics
Numerical Results for 120-node Problem

1000-iteration GA takes roughly the same amount of time as the IP heuristic
Topology Design

The TD Problem
Considerations

- The objective is to find a catalog topology ($Y$) and item placement ($X$) that would minimize the sum of click count cost ($F(Y)$) and page cluttering cost ($C(Y)$).
- Assume every page can be reached from all other pages.
- For a feasible topology, find the best item assignment as in FT.
Two-stage Optimization

\[
\min \left\{ C(Y) + F(Y) : NV^k = b^k, \forall k, \sum_{k=1}^{n} V^k \leq n(n-1)Y, V^k \geq 0, Y_{kl} \in \{0, 1\}, \forall k, l \right\},
\]

where

\[
F(Y) := \min \left\{ \text{trace} \left( FXD(Y)X^T \right) : Xe = e, X^Te = e, X_{il} \in \{0, 1\}, \forall i, l \right\}.
\]

\[
D(Y) := \theta D^B(Y) + (1 - \theta)D^W(Y),
\]

\[
C(Y) = c \sum_{k=1}^{n} \sum_{l=1}^{n} Y_{kl}.
\]
GA Results

- 20 chromosomes, 100 gen. for both FT and TD GAs \((GA + \text{local search})\)
- Finding a good catalog topology is important!

<table>
<thead>
<tr>
<th>(\theta)</th>
<th>Cluttering cost (c = 1)</th>
<th>Cluttering cost (c = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>35.615</td>
<td>21.892</td>
</tr>
<tr>
<td>0.9</td>
<td>27.464</td>
<td>15.055</td>
</tr>
<tr>
<td>0.95</td>
<td>30.847</td>
<td>13.842</td>
</tr>
</tbody>
</table>
The Effect of Page Cluttering

(a) $c = 0.1$

(b) $c = 10$
The Effect of User Demographics

- For $c=1$
- Simplistic designs are suitable for extreme cases

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>Number of links</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>11</td>
</tr>
<tr>
<td>0.1</td>
<td>14</td>
</tr>
<tr>
<td>0.5</td>
<td>13</td>
</tr>
<tr>
<td>0.9</td>
<td>11</td>
</tr>
<tr>
<td>0.95</td>
<td>11</td>
</tr>
</tbody>
</table>
Generic vs. Customized Design
Generic Approach

- Estimating current users’ experience level

\[
\min \left\{ \| D(\theta) - \hat{D} \| : 0 \leq \theta \leq 1 \right\}.
\]

\[
\lambda := \frac{\text{trace} \left( (D^B - D^W)^T (D^W - \hat{D}) \right)}{\| D^B - D^W \|^2} \in [0, 1[,
\]

then set \( \theta^* = \lambda \). Otherwise set \( \theta^* = \begin{cases} 0 & \text{if } \| D^W - \hat{D} \| \leq \| D^B - \hat{D} \|, \\ 1 & \text{if } \| D^W - \hat{D} \| > \| D^B - \hat{D} \|. \end{cases} \)
Customized Approach

- Different designs when the slope changes
- For a 4-node problem, only store 3 best designs out of 912 possibilities (4! times 38 connected topologies)
Concluding Remarks

- Contribution
  - Early work in modeling the catalog navigation design as a math programming problem

- Current and future extensions
  - Clustering of items
  - Cluttering cost?
  - B & C
  - Incorporating product hierarchy, search engine, and hyperlinks
  - Bundling
Questions?
THANK YOU!