The Decision Boundary Map

An Interactive Visual Interface for Making Informed Decisions and Selections in the Presence of Tradeoffs

Klaus Mueller and Shenghui Cheng

Visual Analytics and Imaging Lab
Computer Science Department
Stony Brook University
Decisions are often rendered in the presence of **tradeoffs**
- however, tradeoffs are often difficult to recognize and balance
- especially when there are many factors playing a role

Decisions are often rendered in the presence of **many factors**
- there’s high potential to overwhelm the human decision maker
- they might render decisions that they are not fully sure about
- only to recognize later that better choices could have been made

We propose a **visual interface** to ease this pain
<table>
<thead>
<tr>
<th>Year</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>27</td>
<td>8</td>
</tr>
</tbody>
</table>

Verizon Pharmaceuticals Quarterly Earnings

<table>
<thead>
<tr>
<th>Category</th>
<th>First Quarter</th>
<th>Second Quarter</th>
<th>Third Quarter</th>
<th>Fourth Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
</tr>
<tr>
<td></td>
<td>(Q1)</td>
<td>(Q2)</td>
<td>(Q3)</td>
<td>(Q4)</td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2)</td>
<td>(Q2, Q3)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Q1, Q2, Q3, Q4)</td>
<td>(Q2, Q3, Q4)</td>
<td>(Q3, Q4)</td>
<td></td>
</tr>
</tbody>
</table>

The Data Deluge
We Need Something Intuitive
How About Something Like A Map?

What is your generic term for a sweetened carbonated beverage?

- soda
- pop
- coke
- soft drink

Joshua Katz, Department of Statistics, NC State University
Example: Decide on a College

Many factors (and possible tradeoffs):

- athletics
- academics
- students per faculty
- social life
- tuition
- safety
- weather
- size
- and many more...
no dream school here: good athletics, low tuition, high academic score
Interlude: The Data Matrix

Each college is an N-dimensional vector of attributes

- the data matrix has M rows of colleges each with N attributes
- the M colleges reside in an N-dimensional space of attributes

Goal is to flatten this N-D space into a 2D map with colleges

- this is called embedding
- embedding is essentially an optimization problem
- preserve the N-dimensional distance relations in the 2D layout
Interlude: Space Embedding

General idea:

- preserve N-D space distances $\delta_{ij}$ in 2-D space $d_{ij}$
- minimize

$$stress = \sqrt{\frac{\sum_{ij} (d_{ij} - \delta_{ij})^2}{\sum_{ij} \delta_{ij}^2}}$$

- Multi-Dimensional Scaling (MDS)
- similar data map to similar places
  $\rightarrow$ Similarity Map

Universities
Ivy League
Regional
But...

...are these schools so different?
Embedding the colleges is only half the information
  ▪ we need to embed the attributes as well
  ▪ then we can tell what makes the colleges different
  ▪ it gives the decision context
  ▪ this requires an extension/augmentation to the data matrix

Introducing: the joint data-attribute matrix
The Joint Data-Attribute Matrix

Best of both worlds

- similarity map of the data is based on vector similarity
- similarity map of the attributes is based on pairwise correlation
Achieved by Joint Matrix Optimization

\[
\begin{align*}
D_1 & \quad \cdots \quad D_m \\
V_1 & \quad \cdots \quad V_n
\end{align*}
\]
Optimization schedules for data and attribute layout can vary
- allows us to steer accuracy of the layout to preserve data relationships of attribute relationships or achieve a balance.
MDS Layout Schedules

Glimmer MDS

- Near Set
- Random Set

Data

Data Variables

Variables

Data Variables
MDS Layout Schedules

Glimmer MDS

Near Set

Random Set

Data

Data Variables

Variables

Data Variables
MDS Layout Schedules

Glimmer MDS

- Near Set
- Random Set

Data

Variables

Data Variables

Data Variables

Data Variables
Constructing The Map

Adaptive Kernel Density Estimation

Nadaraya-Watson kernel regression

Contour boundary
Online Deployment

Our application runs interactively in modern web browsers

- data processing on back-end server, interactive graphics on client
- can be accessed anywhere and anytime, free of charge (for now)
- users can simply upload a spreadsheet with the data matrix
- application for a domain name .net and .com is in process
- a rudimentary version is already online and will have been fully developed at the time of the workshop
Use Cases

General public (based on data we have already visualized):

- potential car buyers selecting cars fitting their preferences
- college-age kids and parents selecting the best-fitting college
- wine shoppers navigating the confusing landscape of wines
- investors picking stocks in light of specific investment profiles

Researchers (based on inquiries we have received):

- consulting firms for foundations and nonprofits to inform clients
- disease researchers to gain insight into the interaction of traits
Use Case: Deciding on the Best College

Data are a fusion of two datasets:

- College rankings from US News and World Report
- College Prowler website

A video is online [here](#)

- some annotated stills are presented in the following
The Following Slides are Stills of the Video
Spreadsheet of University Data

Overwhelming and difficult to rank and weigh tradeoffs
The Colleges Shown in Attribute Context

Red nodes: attributes (decision factors); blue nodes: colleges

Similar colleges / attributes map closely

Colleges with higher attribute values map close to that attribute node
Decision Region for Tuition < $27,000

The green regions contain all colleges that fit that condition

tuition slider controls range and shape of the green tuition region (fully accurate)
Add Decision Region for Academic Score $> 8.3$

The purple region contain all colleges that fit that condition.

The academics slider controls the shape of the academics region.

Some colleges fall into both decision regions, tuition and academics: these are the perfect colleges.

Some colleges are close by: these are the trade-off colleges.
Add Decision Region for Athletics Score > 9

The purple region contain all colleges that fit that condition.

Only one college falls into all three decision regions: this is the perfect college for that user (U Virginia).

Some colleges are close by: these are the trade-off colleges.
Use Case: Overviewing Types of Cars

Horsepower: 116~192
MPG: 16~30

European Car

- Euro-Japanese efficient compact cars
- US efficient compact cars
- US semi-efficient medium-power cars
- US big block gas guzzlers
- Euro-Japanese gas guzzlers
- Euro-Japanese semi-efficient medium-power cars
DMN for College Selection Decision

Our framework

- Colleges
- Determine chance of getting in
- Determine fit with academic expectations
- Determine affordability
- Determine fit with social expectations

- Student preferences and academic abilities
- Parents preferences and financial abilities

- Apply to a college

- US News and World Report data
- College Prowler data
User studies suggest that the framework is an intuitive environment to make decisions in selection tasks with many attributes (factors)

Some current work focuses on reducing the number of sliders

- give users options on the attributes they wish to control
- determine a set of good attributes automatically
ANY QUESTIONS?