CrowdSourcing

Top-1 Algorithms over MTurk

Kerui Huang
Department of Computer Science
University of California, Santa Cruz
Image Selection

Top-1 Algorithm
• How well people perform on this kind of task?

• What influences their performance?
Algorithms

The naive algorithm

Another algorithm

Combination Logic
What parameters influence the result?

- Choice of algorithm
- Parameters of the comparison tasks
- Parameters of the User Interface for tasks

Methodology:
Empirical Analysis over MTurk
Outline

• Algorithms
• Experiment Design
• Results
• Conclusions
Outline

• Algorithms
• Experiment Design
• Results
• Conclusions
Algorithms: Tournament

- Parallel
- Less time
Algorithms: Bubble

• Sequential
• More time
<table>
<thead>
<tr>
<th>Tournament</th>
<th>Bubble</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Parallel</td>
<td>• Sequential</td>
</tr>
<tr>
<td>• Less time</td>
<td>• More time</td>
</tr>
</tbody>
</table>
Assumption: Asking for more outputs in a single task leads to a better result, because it is more error-tolerant.
Assumption: Number of input items may influence people’s performance.
Assumption: Number of repetitions per task may influence people’s performance.
Assumption: Number of repetitions for algorithm may influence people’s performance.
Outline

• Algorithms
• Experiment Design
• Results
• Conclusions
Experiment Design

Choose the least blurred image from many blurred images.
• 72 blurred images
• Each image is blurred at different degree.
• The blurriness is in control.

[1] Thanks to Paul Heymann (Bionica) for providing the data set of blurred images.
Experiment Design: UI

1. Number of tasks in a HIT
2. Number of items per task
3. Number of outputs per task
4. Number of assignments
5. How many times the whole algorithm runs
## Experiment Design: Parameters

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
<th>Group 7</th>
<th>Group 8</th>
<th>Group 9</th>
<th>Group 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of questions per task</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Number of answers per task</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Number of tasks per HIT</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Repetitions per Task</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Repetitions for algorithm</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
<td>3</td>
</tr>
</tbody>
</table>
**Experiment Design: Measurement**

\[ error = \sigma_{\text{chosen}_\text{image}} \]

- \( \sigma \) is the standard deviation of Gaussian blur function.
- The blur radius is three times as large as the standard deviation for each image.
- Represent how blurry an image is.
Results: Comparison

<table>
<thead>
<tr>
<th>Group of parameters</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tournament</td>
<td></td>
</tr>
<tr>
<td>Bubble</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of answers in each section</th>
<th>Group 7</th>
<th>Group 2</th>
<th>Group 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
Results: Comparison

<table>
<thead>
<tr>
<th>Group</th>
<th>Group 5</th>
<th>Group 4</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of answers in each section</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Group of parameters

- Error
- Tournament
- Bubble

Group of parameters

- Group 5
- Group 4
- Group 6
### Results: Comparison

<table>
<thead>
<tr>
<th></th>
<th>Group 2</th>
<th>Group 9</th>
<th>Group 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitions per Task</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Repetitions for algorithm</td>
<td>5</td>
<td>15</td>
<td>3</td>
</tr>
</tbody>
</table>

![Bar chart comparing error for different groups and algorithms]
Results: Comparison

- Tournament is more accurate than Bubble.
• Number of input items per task has little influence.

• Number of tasks in a HIT has little influence.
Top-1 Algorithms over MTurk

Kerui Huang
Department of Computer Science
University of California, Santa Cruz

Thank you!