Collecting Social Dynamics on AMT
or
How to become a fake artist?

Subhransu Maji

Social Choice and Networks

Fall 2010
Success and Failure in Cultural Markets

• “winner take all”, “rich getting richer”

• Questions:
  – Is harry potter a really good book?
    • Was rejected by 8 publishers before it sold millions..
  – Friends was rejected by FOX before it went to NBC
  – Beatles struggled to get a record deal...
  – Starwars almost never got made...

• Why:
  – Too many things to consider
  – Social acceptance/commonality
  – Difficulty in predicting success
Possible World Model

• Imagine if we could “rewind” time to when Harry Potter was released and “rerun history” a number of times
• If Harry Potter becomes popular in enough of these worlds then it must be a good book
• Can create such markets artificially on the web

Salganik & Watts, 2009
Salganik and Watts, 2009

- Experiment with music downloads
- About 3,000 users are asked to listen/rate 48 songs from a website
- 8 possible worlds. Total experiment time = 150 days

**Rich-getting-richer dynamics**

**Success is also difficult to predict**
Overview

• Amazon Mechanical Turk’s architecture
• Experimental design
  – Rate abstract art paintings w/ side information
• Results and Conclusions
Amazon Mechanical Turk

Worker:
- Works on tasks
- Gets paid by Amazon

Requester:
- Creates tasks
- Retrieves the answers
- Pays Amazon

HITs = $
HIT : Human Intelligence Task

- “An Amazon Mechanical Turk HIT or Human Intelligence Task is the task you ask a Worker to complete. It may be a task that is inherently difficult for a computer to do.” – AMT Best Practices Guide

All HITs
1-10 of 2165 Results

<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Time Allotted</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPD Inc.</td>
<td>Oct 16, 2010 (6 days 23 hours)</td>
<td>40 minutes</td>
<td>$15.00</td>
<td>1</td>
</tr>
<tr>
<td>TryMyUI Administrator</td>
<td>Oct 29, 2010 (2 weeks 5 days)</td>
<td>2 hours 43 minutes</td>
<td>$10.00</td>
<td>1</td>
</tr>
<tr>
<td>Karen Burnow</td>
<td>Oct 10, 2010 (19 hours 44 minutes)</td>
<td>60 minutes</td>
<td>$6.00</td>
<td>1</td>
</tr>
<tr>
<td>Lawrence Crawford</td>
<td>Nov 22, 2010 (6 weeks 2 days)</td>
<td>10 hours</td>
<td>$5.00</td>
<td>1</td>
</tr>
<tr>
<td>Deborah L Moore</td>
<td>Oct 21, 2010 (1 week 5 days)</td>
<td>30 minutes</td>
<td>$4.50</td>
<td>1</td>
</tr>
<tr>
<td>Still</td>
<td>Oct 16, 2010 (6 days 23 hours)</td>
<td>1 hour 30 minutes</td>
<td>$4.00</td>
<td>1</td>
</tr>
<tr>
<td>Karl Brown</td>
<td>Oct 14, 2010 (5 days 4 hours)</td>
<td>3 hours</td>
<td>$4.00</td>
<td>1</td>
</tr>
</tbody>
</table>
Experimental Design

Rate these paintings on a scale of 1-5.
Use 1 for bad and 5 for excellent.
Click on "Submit Results" once done.

Instructions:
For each painting tell us (1) how much you like it and (2) your guess of the ratings by a professional. You will be paid based on how close your guess of the professional's ratings is to the true answer. For example, do not simply say all the professional's ratings are excellent. You will not be judged based on what your ratings of these paintings are. The descriptions below each of the paintings may be of help.

Side information (fake)

It is not in public display.

your rating: □ 1 □ 2 □ 3 □ 4 □ 5
professional's rating: □ 1 □ 2 □ 3 □ 4 □ 5
Experimental Design

• Artist
  – It is painted by an amateur
  – It is painted by an average artist
  – It is painted by a well known artist
• Location
  – It is not in public display
  – It is at display at a top art museum
  – It is part of someone’s private collection
• Price
  – It is not so expensive
  – It is a moderately expensive painting
  – It is an expensive painting
Results..

- So far 60 different subjects, 360 inputs, 16 hours
- Raw user ratings

(Also same as the order of professional ratings)
Distributions of ratings
Means are similar
Effect of side information

bias = professional’s rating – user’s rating

-0.40(0.98) : It is painted by an amateur
-0.15(1.17) : It is painted by an average artist.
+0.31(1.14) : It is painted by a well known artist.
+0.16(1.09) : It is not in public display.
+0.00(1.09) : It is at display at a top art museum.
-0.08(0.98) : It is part of someone’s private collection.
-0.23(1.10) : It is not so expensive.
+0.20(1.07) : It is a moderately expensive painting.
+0.06(1.16) : It is an expensive painting.
Effect of side information

• +0.31(1.14) : It is painted by a well known artist.

• Stratified by painting:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.56</td>
</tr>
<tr>
<td>0.98</td>
<td>0.98</td>
<td>2.16</td>
<td>0.82</td>
<td>1.38</td>
<td>1.13</td>
</tr>
</tbody>
</table>

More effect for the extremes?
Correlated Results?

- Pearson correlation coefficients between user and professional ratings...
- Seems like side information affects the user ratings!

More effect for the extremes?
Conclusions

• Proof of concept for social experiments on Amazon Mechanical Turk
• Time spend :
  – 1 hour to create the interface
  – 16 hours to collect user input (at the time of creating slides)
  – 2 hours for slides/matlab code to process data
• About 60 unique users/360 ratings..
• Expenses : 3$