Introduction to Micro-task Crowdsourcing Platforms

Lecture 2

Gianluca Demartini

University of Sheffield
Outline

- Micro-tasks
- Micro-task Crowdsourcing platforms
- Amazon MTurk demographics and dynamics
- Examples of micro-task crowdsourcing
### Types of Tasks

<table>
<thead>
<tr>
<th>Task Granularity</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex Tasks</td>
<td>• Build a website&lt;br&gt;• Develop a software system&lt;br&gt;• Overthrow a government?</td>
</tr>
<tr>
<td>Simple Projects</td>
<td>• Design a logo and visual identity&lt;br&gt;• Write a term paper</td>
</tr>
<tr>
<td>Macro Tasks</td>
<td>• Write a restaurant review&lt;br&gt;• Test a new website feature&lt;br&gt;• Identify a galaxy</td>
</tr>
<tr>
<td>Micro Tasks</td>
<td>• Label an image&lt;br&gt;• Verify an address&lt;br&gt;• Simple entity resolution</td>
</tr>
</tbody>
</table>

Example use of micro-task crowdsourcing

- Relevance judgments
- Ontologies
- Museums
- http://www.thesheepmarket.com/

'draw a sheep facing to the left.'
A Crowdsourcing Platform allows **requesters** to publish a crowdsourcing request (**batch**) composed of multiple tasks (**HITs**)

Programmatically Invoke the crowd with APIs or using a website

**Workers** in the crowd complete tasks and obtain a monetary reward
Microtask Aggregators

Enterprise Crowdsourcing Solutions

Crowdflower's technology engages a global workforce to solve your large-scale data problems.

Correct inaccurate business listings.  Assess the relevance of your search results.  Categorize large data sets.  Get quality content in real time.  Tailored solutions to fit your needs.

Samasource.org

1. You send Samasource a project
2. Samasource breaks it down into microwork
3. Work is allocated to our service partners
4. Women, youth, and refugees complete work
5. Samasource compiles work and assures quality
6. Your project gets delivered and helps reduce poverty
Case-Study: Amazon MTurk

- Micro-task crowdsourcing marketplace
- On-demand, scalable, real-time workforce
- Online since 2005 (still in “beta”)
- Currently the most popular platform
- Developer’s API as well as GUI
Amazon MTurk

Make Money by working on HITs

HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:
- Can work from home
- Choose your own work hours
- Get paid for doing good work

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results
Amazon MTurk

- Requesters create tasks (HITs)
- The platform takes a fee (30% of the reward)
- Workers preview, accept, submit HITs
- Requesters approve, download results

- If the results are approved, workers are paid
Demographics of MTurk workers in 2009

Country of residence
- United States: 46.80%
- India: 34.00%
- Miscellaneous: 19.20%

2013 Statistics:
- 1M workers
- 10% active
Demographics of MTurk workers in 2009

Household Income for US workers

Household Income for Indian workers

Year of Birth for US workers

Year of Birth for Indian workers
Demographics of MTurk workers in 2009

http://www.mturk-tracker.com/
For bugs reports or feature requests, please contact Panos Ipeirotis.

If you want to cite this website, please cite the paper Analyzing the Amazon Mechanical Turk Marketplace, P. Ipeirotis, ACM XRDS, Vol 17, Issue 2, Winter 2010, pp 16-21.
MTurk is a Marketplace for HITs

<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date:</th>
<th>Time Allocated:</th>
<th>Reward:</th>
<th>Hits Available:</th>
</tr>
</thead>
<tbody>
<tr>
<td>user</td>
<td>May 23, 2015 (4 weeks 1 day)</td>
<td>25 minutes</td>
<td>$0.05</td>
<td>11526</td>
</tr>
<tr>
<td>user</td>
<td>May 23, 2015 (4 weeks 2 days)</td>
<td>60 minutes</td>
<td>$0.03</td>
<td>23897</td>
</tr>
<tr>
<td>user</td>
<td>May 22, 2015 (4 weeks 1 day)</td>
<td>48 minutes</td>
<td>$0.00</td>
<td>19740</td>
</tr>
<tr>
<td>user</td>
<td>May 22, 2015 (4 weeks 1 day)</td>
<td>60 minutes</td>
<td>$0.00</td>
<td>10734</td>
</tr>
<tr>
<td>user</td>
<td>Apr 30, 2015 (6 days 23 hours)</td>
<td>10 minutes</td>
<td>$0.01</td>
<td>10590</td>
</tr>
<tr>
<td>user</td>
<td>Apr 21, 2016 (51 weeks 6 days)</td>
<td>15 minutes</td>
<td>$0.08</td>
<td>6792</td>
</tr>
<tr>
<td>user</td>
<td>Apr 28, 2015 (5 days 11 hours)</td>
<td>60 minutes</td>
<td>$0.02</td>
<td>6460</td>
</tr>
<tr>
<td>user</td>
<td>May 10, 2015 (2 weeks 2 days)</td>
<td>60 minutes</td>
<td>$0.00</td>
<td>6182</td>
</tr>
<tr>
<td>user</td>
<td>Apr 21, 2016 (51 weeks 6 days)</td>
<td>15 minutes</td>
<td>$0.09</td>
<td>6043</td>
</tr>
<tr>
<td>user</td>
<td>Apr 23, 2015 (8 hours 40 minutes)</td>
<td>60 minutes</td>
<td>$0.10</td>
<td>4882</td>
</tr>
</tbody>
</table>

**Product Attribute Tagging - April 17th** Please read the instructions


**Transcribe up to 25 Seconds of Video to Text - Earn up to $0.12 per HIT**

**Fun and Fast Fashion Tagging**


**Transcribe up to 25 Seconds of General Content to Text - Earn up to $0.14 per HIT**

**What-ame by Gaze "hard mode" | Play a 1min eye tracking game in the web browser" 0416**

**Requester**

**HIT Expiration Date:**

**Time Allocated:**

**Reward:**

**Hits Available:**
# Top requesters

**Top-1000 Requesters, report for May 14, 2015 to June 13, 2015**

<table>
<thead>
<tr>
<th>Requester name</th>
<th>hits</th>
<th>reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turk Experiment</td>
<td>9794</td>
<td>$25,969.90</td>
</tr>
<tr>
<td>video7789</td>
<td>87826</td>
<td>$18,963.40</td>
</tr>
<tr>
<td>Visual Genome</td>
<td>48935</td>
<td>$16,822.55</td>
</tr>
<tr>
<td>VisionTurk</td>
<td>69990</td>
<td>$16,244.14</td>
</tr>
<tr>
<td>Kevin Dodds</td>
<td>139744</td>
<td>$15,928.04</td>
</tr>
<tr>
<td>Jon Brelig</td>
<td>290131</td>
<td>$13,001.36</td>
</tr>
<tr>
<td>Tagasaurus</td>
<td>274148</td>
<td>$7,872.97</td>
</tr>
<tr>
<td>Amazon Requester Inc - browse classification</td>
<td>127864</td>
<td>$7,352.67</td>
</tr>
</tbody>
</table>
Paid Microtask Crowdsourcing scales-out but remains highly unpredictable
SLAs are expensive
mturk-tracker.com

- Collects metadata about each visible batch (Title, description, rewards, required qualifications, HITs available etc)

- Records batch progress (every ~20 minutes)

We note that the tracker reports data periodically only and does not reflect fine-grained information (e.g., real-time variations)

Country-Specific HITs

Workers from US, India and Canada are the most sought after.
Distribution of \textit{Batch Size}
Evolution of Batch Sizes

Very large batches start to appear
HIT Pricing

5-cents is the new 1-cent
Requesters and Reward Evolution

Increasing number of New and Distinct Requesters
HIT Classes

Classify HITs into types (Gadiraju et. al 2014)
- Information Finding (IF)
- Verification and Validation (VV)
- Interpretation and Analysis (IA)
  - Content Creation (CC)
    - Surveys (SU)
  - Content Access (CA)
Supervised Classification
With the Crowd

We trained a Support Vector Machine (SVM) model
- HIT title, description, keywords, reward, date, allocated time, and batch size
- Created labeled data on Mturk for 5,000 HITs uniformly sampled HITs
- Our HIT used 3 repetitions
  - Consensus reached for 89% of the tasks
- 10-fold cross validation
  - Precision of 0.895
  - Recall of 0.899
  - F-Measure of 0.895
- Then, a large-scale classification for all 2.5M HITs
Distribution of HIT Types

Less Content Access batches

Content Creation being the most popular
Batch Throughput Prediction

- Predict batch throughput at time $T$ by training a Random Forest Regression model with samples taken in $[T - \delta, T)$ time span
- 29 Features (including the Type of the Batch)
- Hourly Data in range [June-October] 2014
- We sampled 50 times points for evaluation purposes

We are interested in cases where prediction works reasonably
Predicted vs. Actual Batch Throughput (delta=4 hours)

Prediction Works best for larger batches having large momentum
Significant Features

- What features contribute best when the prediction works reasonably
- We proceed by feature ablation
  - Re-run prediction by removing 1 feature at a time.
  - 1000 samples

\text{HITs\_Available} \ (\text{Number of tasks in the batch})
\text{Age\_Minutes} \ (\text{how long ago the batch was created})
Supply Elasticity

How does the market reacts when new tasks arrive on the platform?
Supply Elasticity

We regressed the percentage of work done (within 1 Hour) against the number of new HITs.
Supply Elasticity

Intercept = 2.5
Slope = 0.5%

20% of new work gets completed within an hour
Summary

• HIT reward has increased over time
• Audio transcription is the most popular task
• Demand for Indian workers has decreased
• Surveys are most popular for US workers
• 1000 new requesters per month join
• 10K new HITs arrive and 7.5K HITs get completed every hour

• Check #mturkdynamics for the main findings
Examples of micro-task crowdsourcing
Information Retrieval

• A document collection
• A user with an information need
• The system has to rank documents by relevance to the user query
IR Evaluation

• Test collections
  – A document collection
  – A set of queries
  – Manual (binary) relevance judgments

• Given the results of a search engine we can
  – Check which results are correct
  – Compute a quality measure
Traditional Collections

• Setup data collection software / harness
• Recruit volunteers (often undergrads)
• Pay a flat fee for experiment or hourly wage
• Characteristics
  – Slow
  – Expensive
  – Tedious
  – Sample Bias
IR Evaluation

- Relevance is hard to evaluate
  - Highly subjective
  - Expensive to measure

- Click data

- Professional editorial work
Why Crowdsourcing for IR?

• Easy, cheap and fast labeling
• Ready-to use infrastructure – MTurk payments, workforce, interface widgets – CrowdFlower quality control mechanisms, etc.
• Allows early, iterative, frequent experiments – Iteratively prototype and test new ideas – Try new tasks, test when you want & as you go
• Proven in major IR shared task evaluations
  – CLEF image, TREC, INEX, WWW/Yahoo SemSearch
Gamification of IR Evaluation

• GeAnn: http://www.geann.org/

• Relevance judgments with Gamification:
  – Text relevance
  – Image relevance

Crowdsourcing Ontology Mapping

• Find a set of mappings between two ontologies

• Micro-tasks:
  – Verify/identify a mapping relationships:
    • Is concept A the same as concept B
    • A is a kind of B
    • B is a kind of A
    • No relation

Crowdsourcing Ontology Mapping

- Crowd-based outperforms purely automatic approaches
Crowdsourcing Ontology Engineering

• Ask the crowd to create/verify subClassOf relations
  – “Car” is a “vehicle”

• Does it work for domain specific ontologies?
  – A “protandrous hermaphroditic organism” is a “sequential hermaphroditic organism”

• Workers perform worse than experts

• Workers presented with concept definitions perform as good as experts

Jonathan Mortensen, Mark A. Musen, Natasha F. Noy: Crowdsourcing the Verification of Relationships in Biomedical Ontologies. AMIA 2013
Summary

• Micro-task crowdsourcing is a growing market
• Multiple platforms available
• Different ways of using it (more about this on Wednesday)