Algorithmic Crowdsourcing
(and Applications in Social Networking)

Jie Wu
Dept. of Computer and Info. Sciences
Temple University
Road Map

- Introduction
- Mechanical Turk
- Applications
- Paradigms
- Challenges and Opportunities
- Social Crowdsourcing
- Conclusion

HotPOST 2013
What
Why
Basic Components

INTRODUCTION
What is Crowdsourcing?

- Coordinating a crowd (a large group of people online) to do micro-work (small jobs) that solves problems (that software or one user cannot easily do)
The Benefits of Crowdsourcing

- Performance
  - Inexpensive
  - Fast

- Human Processing Unit (HPU)
  - More effective than CPU (for some apps)
    - Image labeling
    - Language translation
    - Social network survey
Basic Components

- Requester
  - People submit jobs
  - Human Intelligence Tasks (HITs)

- Worker
  - People work on jobs

- Platform
  - Job management
  - Amazon Mechanical Turk (MTurk)
Worker
HIT
Dashboard

MECHANICAL TURK

HotPOST 2013
Help Find Jim Gray

- Jim Gray, Turing Award winner, went missing with his sailboat outside San Francisco Bay in January 2007.

- Use satellite image to search for his sailboat.
As a worker, make an average of $0.03 per task

Paid directly to Amazon account

As requester, set up simple tasks for workers to complete

Quality control is possible through MTurk services
The Worker

Select a HIT
- By creation date, payment amount, time allotment

HotPOST 2013
Reviewing a HIT

- Review the HIT before accepting
  - Shown full task, allotted time (10 minutes), reward amount ($0.02)

HotPOST 2013
During a HIT

- Shows duration of time
- Gives worker the option to “Return” the HIT

HotPOST 2013
Completing a HIT

- Confirmation message in green
- Automatically shows the next HIT submitted by the same requester
- Check Dashboard to see if HIT is accepted

HotPOST 2013
Sample Dashboard

### Total Earnings

<table>
<thead>
<tr>
<th>Rewards You Have Earned</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved HITs</td>
<td>$4.72</td>
</tr>
<tr>
<td>Bonuses</td>
<td>$0.00</td>
</tr>
<tr>
<td>Total Earnings</td>
<td>$4.72</td>
</tr>
</tbody>
</table>

### Your HIT Status

<table>
<thead>
<tr>
<th>Date</th>
<th>Submitted</th>
<th>Approved</th>
<th>Rejected</th>
<th>Pending</th>
<th>Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$0.00</td>
</tr>
<tr>
<td>Jun 3, 2013</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>$4.72</td>
</tr>
</tbody>
</table>

### HIT Totals

<table>
<thead>
<tr>
<th>HITs You Have Accepted</th>
<th>Value</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HITs Accepted</td>
<td>9</td>
<td>—</td>
</tr>
<tr>
<td>... Submitted</td>
<td>8</td>
<td>88.9%</td>
</tr>
<tr>
<td>... Returned</td>
<td>1</td>
<td>11.1%</td>
</tr>
<tr>
<td>... Abandoned</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HITs You Have Submitted</th>
<th>Value</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HITs Submitted</td>
<td>8</td>
<td>—</td>
</tr>
<tr>
<td>... Approved</td>
<td>7</td>
<td>100.0%</td>
</tr>
<tr>
<td>... Rejected</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>... Pending</td>
<td>1</td>
<td>—</td>
</tr>
</tbody>
</table>
EteRNA

APPLICATIONS: BIOLOGY
EteRNA: CMU, Stanford

- Choose difficulty level
- Solve puzzles: A to U, C to G
- Aim: to gain mastery over the way RNA folds
EteRNA: CMU, Stanford

Welcome to the Strategy Puzzles Series. This series will help you learn basic strategies to solve more advanced puzzles in the Challenges and Lab sections of the game.

"Loops"
A loop is a series of unpaired bases between pairs. One of the loop is shown in the picture below. This loop has 2 loops and it is the 1-1 loop.
EteRNA: CMU, Stanford

MISSION!

Your RNA must fold into the structure in white outline.

You must have 1 or fewer G-C pairs.

You must have 1 or more G-U pairs.

HotPOST 2013
EteRNA: CMU, Stanford
COMMONSENSE KNOWLEDGE
GWAP.com: CMU (no longer available)

ESP Game
- Labeling images

Tag a Tune
- Labeling tunes
Remember this hint:
The keyword 'Grass' is related to 'green', and it is also related to itself, to 'Grass'.

Free Repair

Press ENTER to continue.

Caution, Ensign!
My scans show the next convoy will be accompanied by some basic attack drones.

HotPOST 2013

Press ENTER to continue.
OnToGalaxy: University of Bremen

- Given a keyword
  - e.g., “tourism”

- Collect pods with words related to keyword
  - e.g., “voyage”

- Shoot down pods with unrelated words
  - e.g., “resist”

- An experimental game platform
Galaxy Zoo
Fine-grained Recognition

IMAGE PROCESSING
**GalaxyZoo: Zooniverse**

---

**Classify**

*SHAPE*
Is the galaxy simply smooth and rounded, with no sign of a disk?

- Smooth
- Features or disk
- Star or artifact
GalaxyZoo: Zooniverse

- Zooniverse
  - A website dedicated to citizen science projects

- A platform to label the different galaxies

- Step-by-step instructions and visual guidelines
Sequential
Iterative and Parallel
Divide-and-Conquer
Divide-and-Conquer and Aggregate
Map and Reduce: a Special Case

PARADIGMS
Sequential: Collaborative Workflow

- Lexical translation
  (weak bilinguals or machine)
- Assistive translation
  (strong bilinguals)
- Refine sentence
  (monolinguals)

Iterative and Parallel

- Iterative improve and vote


HotPOST 2013
Divide-and-Conquer

**Step 1: Find**
“Find a close-up image of a flower on flickr.com and load it in GIMP.”

**Step 2: Vote**

**Step 3: Crop**
“Crop the image so that only the petals and top of the stem are visible.”

**Step 4: Vote**

**Step 5: Label**
“Consult Wikipedia to label the petal, pistil, stamen, pedicel, and bract.”

**Step 6: Vote**

**Final**


HotPOST 2013
Divide-and-Conquer and Aggregate

- Divide-and-Conquer and Aggregate
  - Decompose a problem statement and aggregate the results

- Two special aggregates
  - Merge
  - Reduce


HotPOST 2013
Map and Reduce: A Special Case

Challenges
Opportunities

CHALLENGES AND OPPORTUNITIES
Challenges

- **Trade-offs: time, cost, and quality**
  - Max algorithm with human error (with a probability)
  - Maximize quality (via redundancy) subject to cost and time


- **Incentive: money, glory, and love**
  - Platform-centric: a Stackelberg game
  - User-centric: auction-based incentive mechanism

Challenges: HPU + CPU

M. Franklin et al, “CrowdDB: Answering Queries with Crowdsourcing,” SIGMOD 2011

HotPOST 2013
Challenges: Collaborative Workflows

Turkomatic

- Complex works require careful and accurate design workflow
- Problems:
  - Loop subtasks
  - Task starvation

Opportunities

- Beyond simple workflows
  - Graph search
  - Graph match

- Beyond simple worker selection
  - Dynamic procurement

- Beyond independent workers
  - Social networks
Beyond Simple Workflows

- **Graph search**
  - Human-assisted graph search
  - Best sequence of questions with simple Y/N answers

  A. Parameswaran et al, “Human-Assisted Graph Search: It’s Okay to Ask Questions,” VLDB 2010

- **Graph match**
  - People graph (who knows and/or communicates with whom)
  - Puzzle graph (ideas are compatible and can merge)
  - Natural dynamic for people to merge their compatible ideas

Beyond Simple Worker Selection

Dynamic Procurement (multi-armed bandit)

- A gambler facing a row of slot machines
- Which one to play, how many times, and in which order
- Each machine having a random reward from a fixed distribution
- Objective: maximizing the sum of rewards earned through a sequence of lever pulls

Beyond Independent Workers

- Social network of workers
- Iterative recruitment of workers through social ties
- Challenges
  - Graph searching
  - Timeliness of responses
  - Stoppage condition
Computational Surplus Around
AQ Example

SOCIAL CROWDSOURCING
Computational Surplus Around

- Friends help friends
  - Fixed individual capability
  - Probabilistic friends' capability

- Makes dissemination decisions
  - Based on the estimations of the fixed and potential computational capacities

QQ Example

- Tencent QQ, or QQ
  - Instant messaging

- As of March 2013
  - 798.2 million active QQ accounts
  - Peak of 176.4 million simultaneous online users

- QQ experiment
  - Exploring social status of QQ users by responses

HotPOST 2013
Iterative Request/Reply (reduce)

- Initial label is $L = "2"$ (subtract $L$ by 1 when forwarding this request to QQ friends)

- When $L = 0$, return the total number of QQ friends

- When $L > 0$, do the following:
  - Forward this request to all QQ friends
  - After receiving the first 10 replies, compute the average number of friends, and send them back to me
Iterative Request/Reply (merge)

- Initial label is $L = "2"$ (subtract $L$ by 1 when forwarding this request to QQ friends)

- When $L = 0$, return the following:
  - Basic information ($B$)
  - Number of friends ($N$)
  - Timestamps ($T$)

- When $L > 0$, do the following:
  - Forward this request to all QQ friends
  - Pack the first 10 replies, together with your own information ($B, N, T$), and send them back to me

HotPOST 2013
Summary
Acknowledgements

CONCLUSION
Summary

- HPU as a new paradigm to compliment the traditional CPU-based computing

- Many unexplored algorithmic problems
  - Worker selection
  - Social connections of workers
  - Workflow design
  - Cost-time-quality trade-offs
  - Incentive mechanisms
Acknowledgements

- Wei Chang
  Temple Univ.

- Grace Ju
  Carnegie Mellon Univ.