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

Colorado State Univ.
INTRODUCTION

What
Why
Basic Components
Examples

Colorado State Univ.
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

Colorado State Univ.
Basic Components

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

- Worker
  - People work on jobs

- Platform
  - Job management
  - Amazon Mechanical Turk (MTurk)
Example 1: 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.
Example 2: DARPA Network Challenges

- Problem (2009): $40,000 challenge award for the first team to find 10 balloons.
- MIT team won under 9 hours.
- Winning strategy
  - $2,000 per balloon to the first person to send the correct location
  - $1,000 to the person who invited the winner
  - $500 to whoever invited the inviter
  - ... (or to charity) ...
Example 3: Tag Challenges

- Winner from UCSD CrowdScanner: located 3 of the 5 suspects.
- Winning strategy: same as MIT. Also, recruiters of the first 2,000 get $1.
Worker
HIT
Dashboard

MECHANICAL TURK
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
Create a Project: work contents
Create a Project: rewards

- Rewards
- Available work positions
- Deadline
Create a Project: select workers
## Worker: Contract for a HIT

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

Colorado State Univ.
Worker: Avoid Shady Owner

How Turkopticon works:

Turkopticon adds functionality to Amazon Mechanical Turk as you browse for HITs and review status of work you’ve done. As you browse HITs, Turkopticon places a button next to each requester and highlights requesters for whom there are reviews from other workers. Bad reviews let you avoid shady employers and good reviews help you find fair ones. You can view reports made against requesters with a quick click.

As you review HITs you’ve completed, are there HITs you weren’t fairly paid for? Turkopticon adds a button that lets you review requesters from your “Status Detail” page.

Colorado State Univ.
Worker: Reviewing a HIT

Review the HIT before accepting

- Shown full task, allotted time (10 minutes), reward amount ($0.02)

Colorado State Univ.
Worker: During a HIT

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

Colorado State Univ.
Worker: 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**

Colorado State Univ.
Worker: Sample Dashboard

Total Earnings

- Approved HITs: $4.72
- Bonuses: $0.00
- Total Earnings: $4.72

Your HIT Status

Date: Today
- Submitted: 1
- Approved: 0
- Rejected: 0
- Pending: 1
- Earnings: $0.00

Date: Jun 3, 2013
- Submitted: 7
- Approved: 7
- Rejected: 0
- Pending: 0
- Earnings: $4.72

View more...

HIT Totals

- HITs Accepted: 9
  - Value: 9
  - Rate: —

- ... Submitted: 8
  - Value: 8
  - Rate: 88.9%

- ... Returned: 1
  - Value: 1
  - Rate: 11.1%

- ... Abandoned: 0
  - Value: 0
  - Rate: 0.0%

- HITs Submitted: 8
  - ... Approved: 7
  - Rate: 100.0%

  - ... Rejected: 0
  - Rate: 0.0%

  - ... Pending: 1
  - Rate: —
EteRNA

APPLICATIONS: BIOLOGY
EteRNA: CMU, Stanford

- Aim: to gain mastery over the way RNA molecules folds.
By assigning different colors (RNA nucleotides), a RNA chain will fold into different structure
EteRNA: CMU, Stanford

Structure Notation

```
(  )
```

Puzzle Description

"Difficulty Level"
Welcome to the Hairpin RNA Series.

"Science"
A small hairpin RNA or short hairpin RNA (shRNA) is a sequence of RNA that makes a tight hairpin turn that can be used to silence gene expression via RNA interference. shRNA uses a vector introduced into cells and utilizes the U6 or H1 promoter to ensure that the shRNA is always expressed. This vector is usually passed on to daughter cells, allowing the gene silencing to be inherited. The shRNA hairpin structure is cleaved by the cellular machinery into siRNA, which is then bound to the RNA-induced silencing complex (RISC). This complex binds to and cleaves mRNAs which match the siRNA that is bound to it.
**EteRNA**: CMU, Stanford

**MISSION!**

Your RNA must fold into the structure in white outline.

Click here to start!

You must have 2 or more G-U pairs.
EteRNA: CMU, Stanford
EteRNA: CMU, Stanford
GWAP.com
reCAPTCHA
OnToGalaxy

COMMONSENSE KNOWLEDGE
GWAP.com: CMU

ESP Game
- Labeling images

Tag a Tune
- Labeling tunes
OnToGalaxy: University of Bremen

Remember this hint:
The keyword 'Grass' is related to 'green', and it is also related to itself, to 'Grass'.

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?
GalaxyZoo: Zooniverse
GalaxyZoo: Zooniverse
GalaxyZoo: Zooniverse
GalaxyZoo: Zooniverse

- Zooniverse
  - A website dedicated to citizen science projects

- A platform to label the different galaxies

- Step-by-step instructions and visual guidelines
GalaxyZoo: Zooniverse
Fine-Grained Recognition


Colorado State Univ.
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


Colorado State Univ.
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


Colorado State Univ.
Divide-and-Conquer and Aggregate

- Decompose a problem statement and aggregate the results

Two special aggregates
- Merge
- Reduce


Colorado State Univ.
Map and Reduce: A Special Case


Colorado State Univ.
CHALLENGES AND OPPORTUNITIES

Challenges
Opportunities

Colorado State Univ.
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


Colorado State Univ.
Challenges: HPU + CPU

CrowdDB

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

Colorado State Univ.
Challenges: Collaborative Workflows

Turkomatic

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


Colorado State Univ.
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


Colorado State Univ.
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

Colorado State Univ.
Computational Surplus Around
QQ 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

Water Filling Schedule

- Response delay
- Computation (by a friend)
- Reply delay

\[ d_i : \text{response + reply} \]
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

Colorado State Univ.
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

Colorado State Univ.
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 University

- Grace Ju
  Carnegie Mellon University

Colorado State Univ.