Crowdsourcing & Human Computation

Matt Lease
School of Information
University of Texas at Austin
ml@ischool.utexas.edu
“Amazon Remembers”

J. Pontin. *Artificial Intelligence, With Help From the Humans*. NY Times (March 25, 2007)
Amazon Mechanical Turk (MTurk)

• “Micro-task” crowdsourcing marketplace
• On-demand, scalable, real-time workforce
• Online since 2005 (and still in “beta”)
• Programmer’s API & “Dashboard” GUI
From Outsourcing to Crowdsourcing

http://www.mturk-tracker.com (P. Ipeirotis’10)

From 1/09 – 4/10, 7M HITs from 10K requestors worth $500,000 USD (significant under-estimate)
Who are the workers?


• P. Ipeitorotis. March 2010. *The New Demographics of Mechanical Turk*

Road Map

• Example and Introduction
• Crowdsourcing
• Human Computation
• Rethinking Application Design
• The Road Ahead
• Wrap-up
Crowdsourcing
Crowdsourcing

• Take a job traditionally performed by a known agent (often an employee)
• Outsource it to an undefined, generally large group of people via an open call
• New application of principles from open source movement
Other Crowdsourcing Examples

SAILOR MISSING SINCE 1/28/07

Please contact the United States Coast Guard with any information.

Wired Article NY Times Article Ongoing Effort I’d Like to Help! Print a MISSING Poster

Examination Done! We've examined more than 560,000 images from 3 satellites, covering nearly 3,500 square miles of ocean! We currently

INNOCENTIVE

X PRIZE FOUNDATION

Ask 500 People

April 15, 2011

Matt Lease - ml@ischool.utexas.edu
Human Computation
The Mechanical Turk

The original, constructed and unveiled in 1770 by Wolfgang von Kempelen (1734–1804)
The Turing Test (Alan Turing, 1950)
The Turing Test (Alan Turing, 1950)
What is a Computer?

**computer**

[kuh m-pyoo-ter] [Show IPA]

-noun
1. Also called processor, an electronic device designed to accept data, perform prescribed mathematical and logical operations at high speed, and display the results of these operations. Compare analog computer, digital computer.

*Origin:*
1640–50; compute + -er; compare Middle French computeur

-Related forms
computer-like, adjective
non-computer, adjective

Dictionary.com Unabridged
• What was old becomes new

• “Crowdsourcing: A New Branch of Computer Science” (March 29, 2011)

Princeton University Press, 2005
Davis et al. (2010) The HPU.
Human Computation

• People become ‘computists’ once more
  – Do tasks computers cannot (do well)

• 1. Detect robots (Captcha – “reverse Turing test”)”

• 2. Micro-tasks and data labeling (at scale)
  – Game changer for improving practical AI: starving for data

• 3. Rethink what is possible in application design
  – Integrate CPU + HPU = new capabilities
Blending Automation & Human Computation
Soylent: A Word Processor with a Crowd Inside

- Bernstein et al., UIST 2010
Translation by monolingual speakers

- C. Hu, CHI 2009
fold.it

- S. Cooper et al. (2010).
Invisible By-product

CrowdSearch and mCrowd

• T. Yan, MobiSys 2010

mCrowd. Mobile crowdsourcing to sense and protect the planet.
Wisdom of Crowds Computing

Pre-conditions
• Diversity
• Independence
• Decentralization
• Aggregation

**Input:** large, diverse sample
  (increases likelihood of overall pool quality)

**Output:** consensus, selection, distribution
Unreasonable Effectiveness of Data

Massive free Web data changed how we train learning systems

How might access to cheap & plentiful labeled data change the balance again?

April 15, 2011
Matt Lease - ml@ischool.utexas.edu
CrowdForge: MapReduce for Automation + Human Computation

![Diagram showing the process of CrowdForge]

- **Partition**
  - Map
  - Map
  - Reduce
  - Reduce
  - Final

- **Map (facts)**
  - The Empire State Building celebrated its 75th Anniversary on May 1, 2006.
  - ...most popular ones include the Empire State Building, the Statue of Liberty and the Grand Central Terminal.
  - The Statue of Liberty has become an American symbol of freedom and welcome to the immigrants

- **Reduce (paragraph)**
  - Ask most people who plan to travel to New York City what they want to see while they are there and invariably you will hear about the top tourist attractions: the Empire State Building, the Statue of Liberty, and the Grand Central Terminal, with the Empire State Building probably coming in as number one on the list of "must see" for visitors to the city. No wonder: the Empire State Building has a long history, having celebrated its seventy-fifth anniversary on May 1, 2006. Yet the Statue of Liberty is also a popular tourist destination.
Wrap-up
Conclusions

• Shift in practice that’s here to stay
  – Fast, cheap, easy: collect data or do work
  – Emerging phenomenon to study and guide

• New capabilities in application design from automation + human computation

• Hot area, fast changing, many open problems
Thank You!

• Students
  – Catherine Grady (iSchool)
  – Hyunjoon Jung (ECE)
  – Adriana Kovashka (CS)
  – Abhimanu Kumar (CS)

• Omar Alonso, Microsoft Bing

• Support
  – John P. Commons

ir.ischool.utexas.edu/crowd

UT Mechanical Turk & Crowdsourcing Google Group

April 15, 2011  Matt Lease - ml@ischool.utexas.edu
Resources & Upcoming Events

Special issue of Springer’s Information Retrieval journal on Crowdsourcing (papers due May 6, 2011)

Upcoming Conferences & Workshops

• 3rd HCOMP at AAI 2011
• CHI 2011 Workshop (proceedings online)
• SIGIR 2011 Workshop
• TREC 2011 Crowdsourcing Track
• CrowdConf 2011 (TBA)
Data Collection
Snow et al. (2008). EMNLP

• 5 Tasks
  – Affect recognition
  – Word similarity
  – Recognizing textual entailment
  – Event temporal ordering
  – Word sense disambiguation
• 22K labels for $26
• high agreement between Turk annotations and expert “gold” labels
Help Classify Arabic into Dialects!

This task is for Arabic speakers who understand the different local Arabic dialects (اللغات المحلية أو الكرانيات)، and can distinguish them from Fusha Arabic (اللغة الفصحى).

Below, you will see several Arabic sentences. For each sentence:
1. Tell us how much dialect (اللغة) is in the sentence, and then
2. Tell us which Arabic dialect the writer intends.

This following map explains the dialects:

PLEASE READ the following. You MUST understand the classifications, otherwise your work might be rejected!!

- Levantine (اللغة ايلى) does NOT mean 'Syrian' only. It includes Syrian, but ALSO: Jordanian is Levantine, Palestinian is Levantine, and Lebanese is Levantine. That's why all these countries are green in the map.

- Maghrebi (اللغة المغربية) does NOT mean 'Moroccan' only. It includes Moroccan, but ALSO: Algerian is Maghrebi, Tunisian is Maghrebi, and Libyan is Maghrebi. That's why all these countries are purple in the map.

- The word 'dialect' (اللغة) does NOT mean 'spelling mistake' (الخطأ في الإملاء). If the writer was trying to write in 100% correctly, classify it as No dialect, even if it has some spelling mistakes.
Example – Spelling correction
How similar is the artistic style in the paintings above?

- Very similar
- Somewhat similar
- Neither similar nor dissimilar
- Somewhat dissimilar
- Very dissimilar
Models & Incentives

- Pay (e.g. MTurk)
- Fun (or avoid boredom)
- Socialize
- Earn acclaim/prestige
- Altruism
- Learn something new (e.g. English)
- Invisible by-product (e.g. re-Captcha)
- Create self-serving resource (e.g. Wikipedia)

Multiple incentives are often offered in tandem
Altruism

- Contribute knowledge
- Help others (who need knowledge)
- Help workers (e.g. SamaSource)
- Charity
Games with a Purpose (L. von Ahn)

- Players have fun, creators get data as by-product

- distinct from Serious Gaming / Edutainment
  - Player learning / training / education is by-product
Worker Demographics

• 2008-2009 studies found less global and diverse than previously thought
  – US
  – Female
  – Educated
  – Bored
  – Money is secondary
2010 shows increasing diversity

47% US, 34% India, 19% other (P. Ipeitorotis. March 2010)