Lecture #11: Human Computation

Prof. Dr. Sven Seuken
19.4.2012
Housekeeping

• Questions? Concerns?

• Homework assignments
  – MD/CA how due now
  – We will return Bidding Agent hw after class today
  – Amazon Turk hw online by tomorrow
Computation: Now and Then

- “computare” = to count, sum up or reckon together” (cp. early computers)
- Before 1890 “computer” = person who performed calculation as a profession
- People were given a “computing plan” → algorithms

History of Organized Human Computation

- Halley Comet
- The Metric System
- Nautical Almanac
- Prime Meridian
- Coastal Weather Statistics
- The Hollerith Machine
- Telephone Transmission
- Stock Market Prediction
- Agricultural Economics
- Statistical Study of Plants and Animals
- Mathematical Ballistic
- IBM 407 Accounting Machine

(law and von ahn’11)
Human Computation

Definition 11.1 (Human Computation). Given a computational problem from a requester, design a solution using both automated computers and human computers (Law and von Ahn, 2011).

Humans are used as a building block to solve a computational problem: the humans are performing small tasks, yet, within a larger, relatively complex computational system.

- ESP Game
- VizWiz
- re-CAPTCHA
Verify Your Registration

Enter the code shown: 

This helps Yahoo! prevent automated registrations.

Terms of Service

Please review the following terms and indicate your agreement below.

1. ACCEPTANCE OF TERMS
Yahoo! Inc. ("Yahoo!") welcomes you. Yahoo! provides its service to you subject to the following Terms of Service ("TOS"), which may be

By clicking "I Agree" you agree and consent to (a) the Yahoo! Terms of Service and Privacy Policy, and (b) receive required notices from Yahoo! electronically.

I Agree  I Do Not Agree

Verification technology developed in collaboration with Carnegie Mellon University

(von Ahn)
CAPTCHA sweat shops

- Spam companies
- $2.50/hour
- 720 CAPTCHAs / hour
- 1/3 cent / account
“Help! I’ve been waiting for 20 minutes, and nothing happens”
Verify Your Registration

Enter the code shown:  
More info

This helps Yahoo! prevent automated registrations.

Terms of Service

Please review the following terms and indicate your agreement below.  
Printable Version

1. ACCEPTANCE OF TERMS
Yahoo! Inc. ("Yahoo!") welcomes you. Yahoo! provides its service to you subject to the following Terms of Service ("TOS"), which may be

By clicking "I Agree" you agree and consent to (a) the Yahoo! Terms of Service and Privacy Policy, and (b) receive required notices from Yahoo! electronically.

Verifcation technology developed in collaboration with Carnegie Mellon University

(von Ahn)
Free Nude Photos

Type the word in the box if you want to see the next picture

space
Показывать информацию обо мне:
- Всем
- Только зарегистрированным пользователям
- Никому

Защита от автоматической регистрации

\[ \lim_{x \rightarrow 0} \ln \left( 2 + \sqrt{\arctg x \cdot \sin \frac{1}{x}} \right) \]

Введите ответ

X. Очистить

Всё верно

(von Ahn)

Prof. Dr. Sven Seuken - University of Zurich
\[ R_{AB} = \frac{R}{3} \]
What is “1 + 1”?
• 200 million CAPTCHAs a day
• 10 secs per CAPTCHA
• can this be used for good?
reCAPTCHA

- Scan ! OCR ! distort ! reCAPTCHA
- Systems knows one label, not the other
- Show to multiple users, gain confidence in translation (then become known words)
Effectiveness

- facebook, ticketmaster, craigslist, twitter, last.fm,... more than 850m people

(von Ahn)
Application: NY Times

• 1851-1980 archive
• transcribed by 2009 using reCAPTCHA
Type the two words:

bad

Christians

(von Ahn)
Church of Inglip

• http://www.churchofinglip.com/
The Science of Human Computation?

• More art than science?
• How to do this?!

→ Study human computation algorithms!
Human Computation Algorithm

1. input
2. Output

<table>
<thead>
<tr>
<th>Problem</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiplication</td>
<td>set of numbers</td>
<td>a number representing the product</td>
</tr>
<tr>
<td>sorting</td>
<td>set of objects</td>
<td>the set of object re-arranged in a particular order</td>
</tr>
<tr>
<td>medical diagnosis</td>
<td>x-ray and test results</td>
<td>cancer or no cancer</td>
</tr>
<tr>
<td>object recognition</td>
<td>image</td>
<td>a set of tags describing objects in the image</td>
</tr>
<tr>
<td>translation</td>
<td>sentence in language $x$</td>
<td>sentence in language $y$</td>
</tr>
<tr>
<td>editing</td>
<td>text</td>
<td>edited text</td>
</tr>
<tr>
<td>planning</td>
<td>goals and constraints</td>
<td>sequence of actions</td>
</tr>
</tbody>
</table>

3. finiteness (= control, answer after finite #steps)
4. effectiveness (= each operation small enough)
5. definiteness (= each operation unambiguous)

→ reCAPTCHA?

Prof. Dr. Sven Seuken - University of Zurich (law and von ahn’11)
Explicit Control

- **WHAT**: decide what operations need to be performed in what order
- **HOW**: decide how each operation is to be performed
- **WHO**: decide to whom each operation should be assigned

*Human + Machine Intelligence*

(Prof. Dr. Sven Seuken - University of Zurich)
TurkIt scripting language

/* this function creates a recipe (a list of steps) that include the list of ingredients and a recipe name */

function create_recipe(ingredients, recipe_name)
{
    var steps = array();

    // gather a list of steps for making the recipe
    while (!is_done(recipe_name, ingredients, steps)) <-- human-dri:
    {
        new_step = get_next_step(recipe_name, ingredients, steps);
        steps.add(new_step);
    }

    return steps;
}

function is_done(recipe_name, ingredients, steps)
{
    var hitID = once createHIT( ... ‘is_done.html’ ... );
    var answer = once mturk.waitForHIT(hitID)
    return h.answer[0];
}

function get_next_step(recipe_name, ingredients, steps)
{
    var hitID = once createHIT( ... ‘get_next_step.html’ ... );
    var answer = once mturk.waitForHIT(hitID)
    return h.answer[0];
}
Peer Production vs. Human Computation

- Peer Production describes decentralized collaborations among individuals that result in successful large-scale projects, without the use of monetary incentives or managerial hierarchies to organize the group of contributors.

- Features:
  - High modularity
  - Low granularity
  - Low-cost integration helps

1. Wikipedia?
2. Linux?
3. Online Discussion Boards?
4. Stack Overflow?
5. Travel Planner?
6. Yelp!, Tripadvisor?
Linux vs. Human Computation

• Not precise decomposition, “who does what and how” not made specific
• People work on tasks they want and choose → No algorithm for assigning work
• No explicit control over process
• Not definite/finite, no correct solution
• Not small tasks, need lots of advanced knowledge

Prof. Dr. Sven Seuken - University of Zurich
Human Computation vs. Crowdsourcing

• **Crowdsourcing** is a type of outsourcing tasks that are traditionally performed by an employee or contractor, to an undefined, large group of people or community (a crowd) through an open call.

→ Can be used as a tool in human computation systems

• Examples:
  – For HC: Amazon Mechanical Turk (www.mturk.com), Crowdflower, ClickWorker,...
  – Not so good for HC: oDesk, Elance, Freelancer,...
Control Structures

(a) Sequence

(b) Parallel

(c) Selection (Choice)

(d) Repetition (Looping)

(law and von ahn’11)
Challenges

• Correctness: will it solve in a finite number of steps
  – more statistical, adaptive than in standard computation

• Efficiency: how many resources are required?
  – cost of each ‘operation’ a new concern

• Opportunities for optimization
Modeling/Optimization

• Model worker accuracy and voter patterns
• Optimize a plan for how to use works (Dai et al. 2011)
Applications using MTurk

1. Get a second opinion (Mengia)
2. Generate recipes (Balthasar)
3. Translate things in real-time (Andrin)
4. Make subtitles to movies (Malte)
5. Handwriting to Latex (Andras)
6. Traffic jam detection (Khoa)
7. Question answering application (Robert)
8. Finding a similar song (Nico)
9. Find the most interesting sights (Jan)
10. Crowd a travel itinerary (Basil)
Travel Planning: Mobi

• Human Computation?
• Human Computation Algorithm?
  – Explicit control?
  – Structured procedures?
• Peer Production?

→ Crowdware (Zhang et al., 2012)
Soylent: Crowdsourced text editor

- Demo: http://projects.csail.mit.edu/soylent/
- Task Design
  - Granularity
  - Independence
  - Incentives
  - Quality Control

(a) Single, ambiguous operation
(b) Multiple, more well-defined operations
Challenge: Quality

• No incentives for high quality work
• Malicious workers
• Individual differences (e.g., perception on music, summarization task, etc.)

• Ideas
  – output aggregation, redundancy and majority vote (e.g., label)
  – rank aggregation (e.g., rank objects by beauty),.. how to combine?
  – qualification tasks, verification, use of rough machine intelligence, reputation systems
Games with a Purpose (GWAP)

- http://www.gwap.com/gwap/
What do you see?

taboo words

dog
ESP: Incentivize more specific labels

- Generate more taboo words…?
- Two games on top of each other (Mengia)
- Give points based on rarity (Malte)
- Find as many matches as possible per image (Martin)
- Relaxing the time bonus (Nico)
Other Ideas....

- Adjust points based on earlier rounds on this image
- Signal ‘give specific labels’
- Ban gender, color etc. out right
- Automatic synonym matching
<table>
<thead>
<tr>
<th>your descriptions</th>
<th>your partner's descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>male vocal</td>
<td>guitar</td>
</tr>
<tr>
<td>medieval music</td>
<td>solo</td>
</tr>
<tr>
<td>quartet</td>
<td>no vocals</td>
</tr>
<tr>
<td>two females</td>
<td></td>
</tr>
</tbody>
</table>

You: Correct

Score: 80
Timer: 1:41
Bonus: 0

Your partner has chosen.
Peekaboom

PEEK: GUESS WHAT YOUR PARTNER IS REVEALING

BOOM: REVEAL PARTS OF THE IMAGE TO YOUR PARTNER

HINTS HELP YOU GUESS
PASS FOR DIFFICULT IMAGES

GIVE HINTS IF NECESSARY
TELL YOUR PARTNER IF A GUESS IS HOT OR COLD

Prof. Dr. Sven Seuken - University of Zurich

(law and von ahn’11)
Any principles?
Output vs Input agreement

• **ESP game**: see the same input, need to agree on the output

• **Tagatune**: output agreement didn’t work; wasn’t fun. but input agreement (are we listening to the same music?) did work

• **Peekaboom**: asymmetric games, one describer and one guesser; one boomer (image and a word) and one peeker (sees part of image, has to guess secret word)
<table>
<thead>
<tr>
<th>Game</th>
<th>Description</th>
<th>Mechanism</th>
<th>AI Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ESP Game [334]</td>
<td>two players match on a tag for the same image</td>
<td>output-agreement</td>
<td>object recognition</td>
</tr>
<tr>
<td>Peekaboom [337]</td>
<td>player 1 reveals parts of the image associated with a secret word, player 2 must guess the secret word</td>
<td>function computation</td>
<td>object location</td>
</tr>
<tr>
<td>Verbosity [338]</td>
<td>player 1 describes the properties of the entity associated with a secret word, player 2 must guess the secret word</td>
<td>function computation</td>
<td>knowledge extraction</td>
</tr>
<tr>
<td>TagATune [190]</td>
<td>two players exchange tags and determine if two music clips are the same or different</td>
<td>function computation (input agreement)</td>
<td>music classification</td>
</tr>
<tr>
<td>FoldIt [69]</td>
<td>players fold protein structures to minimize total energy</td>
<td>function computation (optimization)</td>
<td>protein folding</td>
</tr>
<tr>
<td>HerdIt [24]</td>
<td>players select tags that describe the music</td>
<td>output-agreement</td>
<td>music classification</td>
</tr>
<tr>
<td>Categorilla [329]</td>
<td>players name an entity that fits a template (e.g., Things that fly)</td>
<td>output-agreement</td>
<td>natural language processing</td>
</tr>
<tr>
<td>MoodSwings [174]</td>
<td>players click on a 2-dimensional grid to indicate the valence and intensity of the mood of a music clip</td>
<td>output-agreement</td>
<td>music mood classification</td>
</tr>
<tr>
<td>Phrase Detective [57]</td>
<td>players identify relationships between words and phrases in a short piece of text</td>
<td>output-agreement</td>
<td>natural language processing</td>
</tr>
<tr>
<td>Phylo [255]</td>
<td>players align colored blocks by moving them horizontally and inserting gaps</td>
<td>output-agreement</td>
<td>genome alignment</td>
</tr>
</tbody>
</table>
Adding a 3rd Player

- ESP game: 3 players that must all agree on the same word → effect?
- ESP game: 3rd player could vote on the resulting word
- ESP game: 3rd player sees four pictures and must decide which one is described
- ESP game: 3 players, the first 2 that match get a point
- ESP game: 3rd player generates taboo words (Mengia)
- Verbosity: 2 guessers, 1 describer (Andrin)
Duolingo

• http://duolingo.com
Bidding Agents: Tournament Result

• 1. Jaanbudget: 56,847.12
• 2. Jebabudget: 55,109.24
• 3. Robibudget: 54,919.50