A HOLISTIC VIEW OF HUMAN FACTORS IN CROWDSOURCING

Sihem Amer-Yahia
CNRS Research Director, France

Keynote at SBBD 2017
October 3rd, 2017
A definition

- The act of posting an open call to hire *cheap, immediate, skilled, and easily accessible* labor online

- A place where one finds work, possibly with *remuneration*

- Micro-tasks often easier to complete by humans than by machines
1. Disaster Management in CrowdMap Ushahidi
2. Audio Transcription
NYPL Lab

Together We Listen
3. Galaxy classification
Galaxy Zoo
4. Receipt Transcription on AMT

Classify Receipt
Hit Reward: $0.02

Real readable original receipt  Not a receipt or not readable

The following details can often be found at the top or bottom of the receipt. Enter as much information as you can find.

Find and enter the business phone number:
Phone
Example: (888) 555-1234 or 8885551234

Find and enter the business address:
Address
City
State          Postal code
Example: 321 Fake Street, Los Angeles, CA, 90210

Next
5. Generic Platforms

Workers

Crowdsourcing Platform

Tasks by requesters

pybossa, foule factory, Prolific, Crowd4U, CrowdFlower, Amazon Mechanical Turk
Many more examples

• **Micro-tasks**
  - audio transcription, text translation, image tagging, citizen science
  - implicit collaboration
  - consensus usually achieved with majority voting

• **Collaborative tasks**
  - a group of individuals *collectively working* to achieve a goal
  - collaborative editing, fan-subbing, solution outsourcing (e.g., Netflix contest)
  - Consensus achieved when crowd converges
Task Deployment Processes and Data

Worker Recruitment

Task Assignment

Task Completion

Result Validation

Worker Compensation

Requester

Workers

Expected wage
Acceptance ratio
Reputation
Skill

Motivation
Incentives
Feedback

Y
Human Factors, a rough characterization

1. Worker-specific
   - *Collaborative tasks*: Affinity, Critical Mass, Interaction model

1. Task-specific
   - *Expected Quality, Budget, Desired Expertise, Incentives*,

2. Workers and tasks
   - *Motivation, Feedback*
Human Factors

- They are pervasive in crowdsourcing processes
- Their acquisition, inference and evolution affect performance
- They have been studied in isolation: one process at a time and one factor at a time
This talk’s purpose and outline

Advocate a holistic approach to human factors
  • because of the unpredictability of humans as resources
  • because human factors are evolving in nature

• Reason 1: Human factors are unpredictable
• Reason 2: Human factors evolve
• Reason 3: Looking beyond
Human Factors Are Unpredictable
Self-appointment to tasks in AMT

<table>
<thead>
<tr>
<th>CTRP: Type name, date and total of a receipt</th>
<th>View a HIT in this group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Requester:</strong> CopyText Inc.</td>
<td><strong>HIT Expiration Date:</strong> Jan 14, 2016 (9 minutes 46 seconds) <strong>Reward:</strong> $0.01</td>
</tr>
<tr>
<td><strong>Time Allotted:</strong> 4 minutes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improve a Transcript</th>
<th>View a HIT in this group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Requester:</strong> CastingWords</td>
<td><strong>HIT Expiration Date:</strong> Jan 14, 2016 (10 minutes 50 seconds) <strong>Reward:</strong> $0.19</td>
</tr>
<tr>
<td><strong>Time Allotted:</strong> 8 hours</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identify groups announced from audio recording (Level 2)</th>
<th>View a HIT in this group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Requester:</strong> TestNotice</td>
<td><strong>HIT Expiration Date:</strong> Jan 14, 2016 (19 minutes 59 seconds) <strong>Reward:</strong> $0.25</td>
</tr>
<tr>
<td><strong>Time Allotted:</strong> 10 minutes</td>
<td></td>
</tr>
</tbody>
</table>
Human Factors in Task Assignment

with Rahman et al. VLDBJ 2015

• **Input**: tasks to complete, human workers
• **Goal**: for one team per task
• **Output**: completed tasks

• Each task has *Expertise, Quality, Budget*
  • *English comprehension* for audio transcription

• Each worker has human factors: *Skill, Expected wage, Acceptance ratio*
Objective: maximize crowd-work quality

Maximize $V = \Sigma_{\forall t \in T} v_t$

$V_t = \begin{cases} 
W_1 \times \Sigma_{\forall j \in \{1..m\}} q_{t,j} + W_2 \times (1 - \frac{w_t}{W_t}) & \text{if } q_{t,j} \geq Q_{t,j} \\
0 & \text{if } q_{t,j} < Q_{t,j} 
\end{cases}$

\[ w_t \leq W_t \]
\[ w_t > W_t \]

where $W_1, W_2 \geq 0$ and $W_1 + W_2 = 1$. 

aggregated worker skills and wages

task quality constraint

task budget
ILP formulation of task quality

\[ q_{t,j} = \sum_{u \in U} u_t \times p_u \times w_{u,j} \geq Q_{t,j}, \forall j \in \{1..m\} \]

\[ w_t = \sum_{u \in U} u_t \times p_u \times w_u \leq W_t \]

\[ u_t = [0/1] \]

\[ X_l \leq \sum_{t \in T} \{u_t\} \leq X_h \]

worker selected or not

worker’s acceptance ratio

worker’s skill

bounds on number of tasks
Task Assignment Solution Overview

NP-hard (reduction using Multiple-Knapsack Problem)

• Our approach: offline indexing for a workload of tasks
  • Our implementation uses the primal-dual barrier method to solve the ILP

• Solution:
  • A greedy randomized algorithm with a $2/5$ approximation factor when objective function is sub-modular
  • A greedy deterministic algorithm with a $1-1/e$ approximation factor when objective function is sub-modular and monotonic
Human Factors are unpredictable

- In practice:
  - Workers may not accept tasks assigned to them
  - Worker and Task churn

- Challenges:
  - How to replace a worker who is not available for a task?
  - How to handle new workers/tasks?

- Our approach: Online index maintenance
  - Solve a marginal ILP problem (on a smaller instance)
Quality Experiments

• **Phase 1**: 8 multi-choice questions/task, to assess skills

• **Phase 2**: Collaborative Document Editing task
  • 20 workers asked to produce reports on 5 different topics:
    1) *Political unrest in Egypt*,
    2) *NSA document leakage*,
    3) *Playstation games*,
    4) *All electric cars*
    5) *Global warming*

• **Phase 3**: Completed tasks evaluated by crowd workers
  • 150 AMT workers evaluated Completeness, Grammar, Neutrality, Clarity, Timeliness, Added-Value
AMT Worker Distributions (Egypt task)

(c) Wage distribution

(d) Strong positive correlation between worker skill and wage
Outcome Quality Assessment

**Playstation Games**

- Completeness: 6
- Added value: 4
- Grammar: 2
- Timeliness: 0
- Neutrality: 2
- Clarity: 4

**Egypt Political Unrest**

- Completeness: 5
- Added value: 4
- Grammar: 3
- Timeliness: 2
- Neutrality: 2
- Clarity: 4
Group-aware Human Factors
with Rahman et al. ICDM 2015

- In some cases, outcome quality was low
  - Conflicting opinions
  - Edit wars

Synergetic effects in working teams.
G. Hertel and G. Hertel.
Journal of Managerial Psychology 2011
Affinity
- Type Indicator: MBTI. Myers and Briggs. Consulting Psychologists Press 1988
- Are two heads better than one? Crowdsourced translation via a two-step collaboration of non-professional translators and editors. R. Yan et. al. ACL 2014

- Intra-team distance: e.g., edit wars

\[ \text{DiaDist}(G) = \max_{\forall u_i, u_j \in G} \text{dist}(u_i, u_j) \]

Critical Mass
Managing research quality: critical mass and optimal academic research group size. R. Kenna et. al. IMA Journal of Management Mathematics 2012
Objective, revisited

Minimize \( \{ \text{DiaDist}(G) + \sum_{G_i, G_j \in g} \text{SumInterDist}(G_i, G_j) \} \)

- Under:

\[
\sum_{u_i \in G} u_{d_i} \geq Q_i \quad \forall d_i \\
\sum_{u \in G} w_u \leq C \\
|G_i| \leq K \quad \forall i = \{1, 2, \ldots, x\}
\]

skill

\[
\sum_{u_i \in G} u_{d_i} \geq Q_i \quad \forall d_i
\]

cost

Critical mass

\[
\sum_{u \in G} w_u \leq C
\]
Overview of Algorithmic Solutions

• A two-stage approach
  1. Form a single team that maximizes intra-affinity, and satisfies skill and cost (NP-hard, reduction of Min-Dia, a variant of Compact Location)
  2. Decompose into smaller teams, each satisfies critical mass and inter-affinity across teams is maximized (NP-hard, reduction of Minimum Bisection)

• Algorithms
  1. An instance optimal exact algorithm and a 2-approximation algorithm (when distance is a metric)
  2. A 3-approximation algorithm (akin to Min k-cut)
Experiments with Affinity and Critical Mass

- Translation task with 120 AMT workers
- Region- and age/gender-based affinities

Results
- Higher affinity impacts positively quality
- A group beyond size 10 is less effective
- Region more effective than age/gender
Human Factors Evolve
In practice...

- Workers are involved in a series of tasks
- Their motivation evolves over time
Motivation in the Social Sciences

Motivation through the design of work: Test of a theory. J. Hackman and G. R. Oldham. Organizational behavior and human performance, 1976

• 658 employees in 62 heterogeneous jobs (white collar, blue collar, industry, services, urban and rural settings) in 7 organizations.

• Goal: which Job Dimensions stimulate which Psychological States: experienced meaningfulness of the work, experienced responsibility for the work outcomes, knowledge of the actual results of the work.

• Proposed model good for job design, i.e.
  • in determining the potential of a job to engender motivation,
  • in identifying which jobs need improvement,
  • in assessing the readiness of employees to respond to a redesigned job
Motivation in the Social Sciences


\[
\text{Motivating Potential} = \left[ \frac{\text{Skill}}{\text{Variety}} + \frac{\text{Task}}{\text{Identity}} + \frac{\text{Task}}{\text{Significance}} \right] \times \text{Autonomy} \times \text{Feedback}
\]
What is motivation (in AMT)?

More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. N. Kaufmann, T. Schulze, and D. Veit. *AMCIS 2011*

<table>
<thead>
<tr>
<th>Enjoyment Based Motivation</th>
<th>Skill Variety</th>
<th>Task Identity</th>
<th>Task Autonomy</th>
<th>Direct Job Feedback</th>
<th>Pastime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Community Based Motivation</th>
<th>Community Identity</th>
<th>Social Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Immediate Payoffs</th>
<th>Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delayed Payoffs</th>
<th>Signaling</th>
<th>Human Capital Advancement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social Motivation</th>
<th>Action Significance by Values</th>
<th>Action Significance by Norms &amp; Obligations</th>
<th>Indirect Job Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

|                   |                               |                                             | 1.7                   |
In the related work

• Incentives
  • with gradually increasing pay (Gao et al, PVLDB 2014)
  • with a bonus (Shaw et al. CSCW 2011, Slivkins et al. WWW 2015)
  • with feedback on others’ performance (Shaw et al. CSCW 2011)
  • with entertainment during task completion (Dai et al, CSCW 2015)

• Feedback
  • CrowdFlower displays a panel above the task with the worker’s estimated accuracy so far
  • Encouraging text and a board showing how well a worker is doing. Diversions improve worker retention rate while retaining the same work quality (Rzeszotarski et al, AAA 2013)
Proposed Approach

Observe workers and adaptively assign tasks that maximize their motivation.
Focus on two Factors

Enjoyment Based Motivation
- Skill Variety: 2.4
- Task Identity: 2.3
- Task Autonomy: 2.4
- Direct Job Feedback: 2.0
- Pastime: 2.1

Community Based Motivation
- Community Identity: 2.0
- Social Contact: 1.3

Immediate Payoffs
- Payment: 3.0

Delayed Payoffs
- Signaling: 1.9
- Human Capital Advancement: 2.2

Social Motivation
- Action Significance by Values: 1.7
- Action Significance by Norms & Obligations: 1.0
- Indirect Job Feedback: 1.7
Adaptive Task Assignment (motivation)

with J. Pilourdault, S. B. Roy, D. Lee. EDBT 2017

balance between 2 factors, e.g.,
intrinsic factor, task diversity, and
extrinsic factor, task reward

\[ \text{motiv}(\mathcal{T}, w) = \alpha_w \cdot TD(\mathcal{T}) \]
\[ + \beta_w \times TR(\mathcal{T}, w) \]
Adaptive Task Assignment (optimization)

For a worker, find a set of tasks:

\[
\begin{align*}
\arg \max \quad & \sum_{w \in \mathcal{W}^i} \text{motiv}(\mathcal{T}^i_w, w) \\
\forall w \in \mathcal{W}^i, \quad & |\mathcal{T}^i_w| \leq X_{max} \quad (C_1) \\
\forall w, w' \in \mathcal{W}^i, \quad & \mathcal{T}^i_w \cap \mathcal{T}^i_{w'} = \emptyset \quad (C_2)
\end{align*}
\]
Adaptive Task Assignment (performance)

• 158,018 tasks from CrowdFlower in 22 kinds

• 23 workers in AMT

• 2 task assignment strategies:
  • Tasks matching a worker’s profile: RELEVANCE
  • Tasks achieving a diversity/payment balance: DIV-PAY
Task Throughput

![Bar chart](chart.png)

- **Relevance**: 2.3 completed tasks/min
- **Div-Pay**: 1.5 completed tasks/min
Worker Retention

% of sessions that ended after $x$ tasks were completed

#completed tasks
Outcome Quality

- Relevance: 67%
- Div-Pay: 73%

% correct answers
Summary and Takeaways

- Human factors are essential in crowdsourcing
- They need to be observed during task completion and leveraged in task assignment
- Their evolving nature requires to optimize crowdsourcing processes holistically
Human Factors in Task Assignment

1. Worker Recruitment
2. Task Assignment
3. Task Completion
4. Result Validation
5. Worker Compensation

Y
Adaptive Task Assignment

1. Worker Recruitment
2. Task Assignment
3. Task Completion
4. Result Validation
5. Worker Compensation

Y
Learning Human Factors

1. Worker Recruitment
2. Task Assignment
3. Task Completion
4. Result Validation
5. Worker Compensation
Looking Beyond
Toward Holistic Optimization

1. Worker Recruitment
2. Task Assignment
3. Task Completion
4. Result Validation
5. Worker Compensation

Flowchart showing the process of向 Towards Holistic Optimization with steps 1 to 5.
Why does a holistic view matter that much?

*More than fun and money. worker motivation in crowdsourcing—a study on mechanical turk. N. Kaufmann, T. Schulze, and D. Veit. AMCIS 2011*

---

**Enjoyment Based Motivation**
- Skill Variety: 2.4
- Task Identity: 2.3
- Task Autonomy: 2.4
- Direct Job Feedback: 2.0
- Pastime: 2.1

**Community Based Motivation**
- Community Identity: 2.0
- Social Contact: 1.3

**Immediate Payoffs**
- Payment: 3.0

**Delayed Payoffs**
- Signaling: 1.9
- Human Capital Advancement: 2.2

**Social Motivation**
- Action Significance by Values: 1.7
- Action Significance by Norms & Obligations: 1.0
- Indirect Job Feedback: 1.7
Crowdsourcing platforms as a learning destination

• Ability for workers to express:
  1. I have x amount of time, which tasks should I complete?
  2. I want to make x dollars, which tasks should I complete?
  3. *I want to improve some skill, which tasks are best suited to me?*

• Requires to capture human factor end-to-end
  • to model team formation (who to team up with to learn faster?)
  • to model feedback in result validation and include it in task assignment