Strategies for integrating people and machine learning in online systems

Jason Laska Ph.D.
Machine Learning @ Clara Labs
June 2017

presenting the work of
Michael Akilian, Briana Burgess, Joey Carmello, Matthew Ebeweber, David Gouldin, Evan Hadfield, Olga Narvskaia, Maran Nelson, Jodi Nicolli, Emily Pitts, Gavin Schulz, Oliver Song

www.claralabs.com  @chappaquack  @claralabs
Email Coordination

Hi Elie,
Let’s try to meet next week and discuss a few ideas for the new marketing project.
I have added Clara, who find us a good time.

Helen

Hi Elie,
Happy to get something on the calendar for you and Helen.

How does 10am PDT or 2pm PDT on Wednesday, June 25th work for you?

Best,
Clara

Clara,
1:30 on Wed 6/25 should work, can call me at 415-314-1519
Sorry for the delay- traveling and only on mobile for the week
Sent from my iPhone

Hi Elie,
Great, I just sent an invite with the event details.

Have a great Friday!

Best,
Clara
Email Coordination

Hi Ellie,
Let’s try to meet next week and discuss a few ideas for the new marketing project. I have added Clara, who finds us a good time.

Helen

Hi Ellie,

Happy to get something on the calendar for you and Helen.

How does 10am PDT or 2pm PDT on Wednesday, June 25th work for you?

Best,
Clara

Clara,

1:30 on Wed 6/25 should work, can call me at 415-314-1519

Sorry for the delay—traveling and only on mobile for the week

Sent from my iPhone

Hi Ellie,

Great, I just sent an invite with the event details.

Have a great Friday!

Best,
Clara

Customer Preferences

Scheduling assistant

FOLLOW-UP LIMIT
How many times should Clara follow up?

3

DAILY MEETING CAP
How many meetings do you want to have per day?

3

Locations and preferences that Clara will use to schedule in-person meetings.

COFFEE (DEFAULT)
Red Door Coffee

111 Minna St, San Francisco, CA 94105, USA

OFFICE
Clara Labs

576 Sacramento St, San Francisco, CA 94111, USA
“Let’s meet in Greenwich for coffee next week.”
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

**Location:** Greenwich, CT
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

**Location:** Greenwich, CT

**Coffee:** 8am — Noon (preference)
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

**Location:** Greenwich, CT  
**Coffee:** 8am — Noon  
**Max Daily Meetings:** 3

<table>
<thead>
<tr>
<th>Time</th>
<th>M</th>
<th>Tu</th>
<th>W</th>
<th>Th</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>8a</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
</tr>
<tr>
<td>12</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>OOO</td>
<td>Lunch</td>
</tr>
<tr>
<td>5p</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
<td>busy</td>
</tr>
</tbody>
</table>

**recurring location:** NY CT CT NY
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

Location: Greenwich, CT
Coffee: 8am — Noon (preference)
Max Daily Meetings: 3 (preference)

Apply NLP on calendar:

OOO: out of the office, what does that mean in this context?
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:
- **Location**: Greenwich, CT
- **Coffee**: 8am — Noon (preference)
- **Max Daily Meetings**: 3 (preference)

Apply NLP on calendar:
- **OOO**: out of the office, what does that mean in this context?
- **Lunch**: can we schedule over this or is it important?
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

**Location:** Greenwich, CT

**Coffee:** 8am — Noon (preference)

**Max Daily Meetings:** 3 (preference)

Apply NLP on calendar:

**OOO:** out of the office, what does that mean in this context?

**Lunch:** can we schedule over this or is it important?

**Relax:** can relax constraints if there’s enough travel time?
“Let’s meet in Greenwich for coffee next week.”

Apply constraints:

- **Location**: Greenwich, CT
- **Coffee**: 8am — Noon (preference)
- **Max Daily Meetings**: 3 (preference)

Apply NLP on calendar:

- **OOO**: out of the office, what does that mean in this context?
- **Lunch**: can we schedule over this or is it important?
- **Relax**: can relax constraints if there’s enough travel time?
**Example: Suggesting times**

**Customers really want**
graceful and intuitive
default-case handling
How Clara handles this example
How Clara handles this example
How Clara handles this example

- Preference constraints
- Participant availabilities/unavailabilities
- Any accessible party calendars

Integrated with calendar
"Let's meet in Greenwich for coffee next week."
**TASK TYPE: Predict & Annotate**

- fix incorrect predictions
- augment with missing parameters

*feedback loop to machine learning*

---

**Available Times**

THURSDAY, 20 APR, 2:00PM - 2:30PM PDT

From email message: "Yes, let’s do Thursday @ 2p.

---

30 minutes 🕒

Duration found in email message: "...to schedule a 30 minute touch base. Looking..."
Breaking work into tasks

**TASK TYPE:** Compute & Review

**state:** new

**location:** Greenwich

**channel:** coffee

**time-pref:** next week

**intent:** schedule

**action:** suggest times
**TASK TYPE:** Compute & Review

- check output makes sense
- check customer needs met

feedback loop to product and engineering
TASK TYPE: Manual Override

* manual scheduling
* highly credentialed worker
Automation is a spectrum

tasks fully to partially automated

leverage task differences
match task difficulty with processing skill (person or machine)

cost and speed gains without full automation

requires

tightly integrate worker-operations, machine learning, and UX-design

avg throughput by hour

<table>
<thead>
<tr>
<th></th>
<th>avg throughput by hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sessions</td>
<td>~1.8x</td>
</tr>
<tr>
<td>sessions with worker</td>
<td>~1.4x</td>
</tr>
</tbody>
</table>

feb | mar | apr
<table>
<thead>
<tr>
<th>constraints (requirements)</th>
<th>challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>high accuracy bar</td>
<td>people are naturally noisy at data entry</td>
</tr>
<tr>
<td>bounded processing cost</td>
<td>people hours are more expensive than cpu cycles/hr</td>
</tr>
<tr>
<td>workforce size</td>
<td>learning the platform</td>
</tr>
<tr>
<td>bounded processing time</td>
<td>people are naturally slow at data entry tasks</td>
</tr>
<tr>
<td>workforce elasticity</td>
<td>“staffing on a dime”</td>
</tr>
</tbody>
</table>
Worker lifecycle

- self-directed program
- build commitment to platform through promotion
Sandbox environment

- Task
  - Replicate state
  - Submit
    - Production
      - Candidate 1 score
      - Candidate 2 score
      - Candidate n score

- Candidates
  - Aggregate & compare

- Sandbox — "scheduling multiverse"
Task assignment

tasks

automatable • • • easiest (most automated) (less automated) hard
Task assignment

- **cost**: $ - $$$
- **tasks**: automatable → easiest (most automated) → hard (less automated)
- **queue**: beginner → expert
- **worker pool**: automatable tasks (most automated) → easiest tasks (least automated) → hard tasks
Task assignment

- Automatable tasks (most automated)
- Easiest tasks (less automated)
- Hard tasks

- Escalate

- Cost: $ $ $$$

- Worker pool

- Queue
Task assignment

- cost
  - $ (automatable)
  - $$$ (easiest, most automated)
  - $$$ (hard, less automated)

- tasks
  - automateable → easiest → escalate → hardest

- queue
  - beginner

- worker pool
  - expert
Work recycling

recycle the first queue n times
- beginners may escalate when unsure (equivalent to “skip” in this case)
- # of recycles = proxy to task difficulty
- works well with incentive to avoid mistakes
Work recycling

Recycle the first queue $n$ times
- Beginners may escalate when unsure (equivalent to “skip” in this case)
- # of recycles = proxy to task difficulty
- Works well with incentive to avoid mistakes

Other good strategies

“Double or Nothing: Multiplicative Incentive Mechanisms for Crowdsourcing” by N. Shah, D. Zhou

- Allow worker to “skip” tasks
- Reward based on known-examples:
  - No pay for skipping
  - High penalty for incorrect labels
  - High reward for correct labels

Incentivizes skipping if worker confidence is low
<table>
<thead>
<tr>
<th>constraints (requirements)</th>
<th>challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>high accuracy bar</td>
<td>people are naturally noisy at data entry</td>
</tr>
<tr>
<td>bounded processing cost</td>
<td>people hours are more expensive than cpu cycles/hr</td>
</tr>
<tr>
<td>workforce size</td>
<td>learning the platform</td>
</tr>
<tr>
<td>bounded processing time</td>
<td>people are naturally slow at data entry tasks</td>
</tr>
<tr>
<td>workforce elasticity</td>
<td>“staffing on a dime”</td>
</tr>
</tbody>
</table>
Incentives

### Competing Incentives

<table>
<thead>
<tr>
<th></th>
<th>Worker Throughput</th>
<th>Worker Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per-task payment</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>Time preference (by ranked accuracy*)</td>
<td>↓</td>
<td>↑</td>
</tr>
</tbody>
</table>

*Workers must also maintain a minimum accuracy to remain on the platform.

*Competing incentives drive both throughput and accuracy.*
Staffing: Supply and demand

Jodi’s 3-pass assignment algorithm

- predicted demand/hour
- worker reliability
- worker throughput
- worker accuracy

requested hours (all potential weekly supply) → hour assignments

when supply > demand: 😞

when supply << demand: workers find out about “pick up” work via turtlebot in community Slack channel

ExtraHelpBot 5:50 AM
@here Extra help requested for the 6 AM PDT hour - 1 CRAs please.
Measuring accuracy: Mistake tracking

- worker peer review
- workers have access to previous annotations
- workers have task context
- function of time/cost constraints
Measuring accuracy: Mistake tracking

worker peer review
- workers have access to previous annotations
- workers have task context
- function of time/cost constraints

other good strategies
1) assign multiple workers to same task
2) estimate underlying annotation values and worker accuracy and biases
   a) jointly
      e.g., “Quality Management on Amazon Mechanical Turk” by Ipeirotis, Provost, & Wang
   b) via Cohen/Fleiss kappa: $\kappa := 1 - \frac{1 - p_{\text{observed agreement}}}{1 - p_{\text{chance agreement}}}$
Community is essential

“Clara Remote Assistant” (CRA)

**community manager**
- runs worker support
- builds platform documentation
- community pulse surveys
- promote the needs of the community

**beta launches**
- CRAs love trying new tools before they roll out
- we get amazing feedback

**community Slack channel**
- ask/answer questions, pictures of pets, turtlebot, etc.

we love the CRAs!!
**design problems**
combining ML & UX for data-entry systems

**software problems**
modeling email requests defining new Clara capabilities
worker task management

**ML problems**
natural language understanding
continuous model train and integration

**kinds of people we’re looking for**
fullstack & frontend engineers
machine learning engineers

*we work in python/flask, react/redux, aws services, sklearn, keras…*

www.claralabs.com  
@chappaquack  
@claralabs

Come talk to me!

jason@claralabs.com

Come work with us!