Programming with People

Emery Berger
University of Massachusetts Amherst
“is this a giraffe?”
isGiraffe( )
isGiraffe()
isGiraffe()
isGiraffe( )
isGiraffe()
Please enter the name of the **product** being advertised

- This is not an ad
- The image failed to load
- The Product or Company Name is not known
- Contains adult content
isGiraffe()
isGiraffe() -> True
isGiraffe() => False
False
random adversary model
random adversary model
random adversary model
random adversary model
Qualifications Required:
HIT abandonment rate (%) is less than 20
HIT approval rate (%) is greater than 95

Qualifications Required:
Total approved HITs is greater than 1000
Categorization
HIT approval rate (%) is not less than 95

Qualifications Required:
Total approved HITs is greater than 2500
HIT approval rate (%) is greater than 95
Location is US

Long-term financial incentive
credentials limit *Sybil* attacks

**Step 3:** Get paid for your work

After the requester approves your work, money is deposited into your Amazon Payments account.
isGiraffe( )

?
Pr[agree] = \frac{1}{2}
isGiraffe( )
Pr[agree] = 1/32 < 5%
\[ \Pr[\text{agree}] = k \left( \frac{1}{k} \right)^n \]

(unanimous agreement)
isGiraffe( )
isGiraffe(scala)
AutoMan programmer-specified

Total $ for computation
isGiraffe() (Scala)

AutoMan programmer-specified

Total $ for computation

95% (p < 0.05) Confidence level (per function)
isGiraffe()
isGiraffe()
isGiraffe()

US minimum wage, Adaptive doubling prevents gaming

30s, $0.06 ($7.25 / 120)
60s, $0.12
$E[\text{gain}] = \text{base} \ (P_{\text{avail}})^{\text{round}} \ * \ \text{multiplier}^{\text{round}}$
$ E[\text{gain}] = \text{base} \times (\frac{1}{2})^{\text{round}} \times \text{multiplier}^{\text{round}} $
\[ E[\text{gain}] = \text{base} \ (\frac{1}{2})^{\text{round}} \times 2^{\text{round}} \]
E[gain] = base
\[ E[\text{gain}] = \text{base} \]

no incentive to wait
isGiraffe( )
isGiraffe() = True, 95% confidence
How many giraffes are in this picture?

- None
- More than one
- One

k choices
How many giraffes are in this picture?

- None
- More than one
- One

k choices
How many giraffes are in this picture?

- None
- More than one
- One

k choices
How many giraffes are in this picture?

- None
- More than one
- One

k choices
How many giraffes are in this picture?

- More than one
- None
- One

k choices
Which are from Sesame Street?

☐ Oscar the Grouch
☐ Kermit the Frog
☐ Spongebob Squarepants
☐ Cookie Monster
☐ The Count

$2^k$ choices
Which are from Sesame Street?

☑ Oscar the Grouch
☑ Kermit the Frog
☑ Spongebob Squarepants
☑ Cookie Monster
☑ The Count

$2^k$ choices
Which are from Sesame Street?

- Oscar the Grouch
- Spongebob Squarepants
- Cookie Monster
- The Count

2^k choices
What does this license plate say?

What characters are printed on this license plate?

\[ [A-Z0-9]^6 \]

\[ 36^6 \text{ choices} = 2176782336 \]
Which one of these doesn’t belong?

[95% conf.]

**AUTO MAN:** spawns 3 tasks @ $0.06; 30s work

\[
\begin{align*}
& t_1 & t_2 & t_3 \\
\end{align*}
\]
Which one of these doesn’t belong? [95% conf.]

**AUTO MAN**: spawns 3 tasks @ $0.06; 30s work

\[t_1 \quad t_2 \quad t_3\]

\[1m 50s\]
Which one of these doesn’t belong?

[95% conf.]

**AUTOMan:** spawns 3 tasks @ $0.06; 30s work

- $t_1$
- $t_2$
- $t_3$

- 1m 50s
- 2m 30s
Which one of these doesn’t belong? [95% conf.]

AUTOman: spawns 3 tasks @ $0.06; 30s work

$1m 50s$

$2m 30s$

$2m 50s$
Which one of these doesn’t belong?
[95% conf.]

**AUTOMAN:** spawns 3 tasks @ $0.06; 30s work

\[ t_1 \quad t_2 \quad t_3 \]

1m 50s

2m 30s

2m 50s

Inconclusive!
Which one of these doesn’t belong?
[95% conf.]

AutoMan: spawns 3 more tasks
Which one of these doesn’t belong? [95% conf.]

**AUTO MAN:** spawns 3 more tasks

\[ t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \quad t_6 \]

7m
Which one of these doesn’t belong? [95% conf.]

AUTOMAN: spawns 3 more tasks

\[ t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \quad t_6 \]

7m
18m 50s
Which one of these doesn’t belong? [95% conf.]

**AUTOMan: spawns 3 more tasks**

- $t_1$
- $t_2$
- $t_3$
- $t_4$
- $t_5$
- $t_6$

7m
18m 50s
51m
Which one of these doesn’t belong? [95% conf.]

AUTOMAN: spawn 1 more task @ $0.12, 60s work
Which one of these doesn’t belong?

[95% conf.] 

**AUTOMan**: spawn 1 more task @ $0.12, 60s work

1h 9m 50s; cost = $0.36

**AUTOMan**: 5 out of 6

⇒ 95% confidence;

return
read_plate(  )
Recognize this license plate.

Requester: Dan Barlow
Qualifications Required: None

What are the numbers printed on this license plate? (CAPS ONLY, NO DASHES, DOTS OR SPACES, please!)
def is_car(img_url: String) =
  a.RadioButtonItemQuestion { q =>
    q.budget = 1.00
    q.confidence = 0.95
    q.text = "Is this a car?"
    q.image_url = img_url
    q.options = List(
      a.Option('yes, "Yes"),
      a.Option('no, "No"
    )
  )
}
def get_plate_text(img_url: String) =
a.FreeTextQuestion { q =>
  q.text = "What does this plate say?"
  q.image_url = img_url
  q.pattern = "XXXXXXYYYY"
}
val plate_texts = s3_urls.par.map
{ url => get_plate_text(url) }
plate_texts.foreach { text => println(text) }

def get_plate_text(img_url: String) =
a.FreeTextQuestion { q =>
  q.text = "What does this plate say?"
  q.image_url = img_url
  q.pattern = "XXXXXYYYY"
"Is this a vehicle?"

yes

"What does the license plate say?"

767JKF
MediaLab LPR database

“extremely difficult” dataset

144 plates

Accuracy: 91.6%

Average cost: 12.08 cents

Latency: < 2 minutes per image
How many giraffes are in this picture?

- None
- More than one
- One

Which are from Sesame Street?

- [x] Oscar the Grouch
- [] Kermit the Frog
- [x] Spongebob Squarepants
- [x] Cookie Monster
- [] The Count

read_plate(767-JKF)
3.76 mi  1:46:40 min  28'20" pace

CALORIES BURNED
835 cals

IMPACT ON YOUR DAY
+11,599 of 14,529 steps taken
+835 of 2,717 calories burned
1. take a photo  
2. algorithms (???)  
3. return estimate

230.3 kcal
1. take a photo
2. algorithms (913g)
3. return estimate

230.3 kcal
Im2Calories: towards an automated mobile vision food diary

Once we have determined that the image contains a meal, we try to analyze its contents. The first step is to determine which restaurant the user is in. In this paper, we use Google’s Places API [27] for this. We then retrieve the menu of the nearest restaurant from the web,\(^5\) parse the menu list of \(K\) food items, and retrieve images for each item.
1. take a photo  
2. machine learning  
3. return estimate  

230.3 kcal
1. take a photo
2. crowdsourcing
3. return estimate

230.3 kcal
readPlate()
“How many calories are in this donut?”
1. take a photo  
2. crowdsourcing  
3. return estimate
VoxPL

Extends AutoMan DSL with \texttt{estimates}

```scala
def numCalories(url: String) = Estimate(
  confidenceInterval = SymmetricCI(50),
  text = "How many calories are in the food pictured?",
  imageUrl = url
)
```
“How many calories are in this donut?”

2, 326, 214, 318, 283, 274

$214 \text{ median} \neq 288.\overset{\text{?}}{5} 274$
What would it take to trust that median value of 278.5 is a good estimate?

Donut contains 278.5±50 kcal.
What would it take to trust that median value of 278.5 is a good estimate?

“Donut contains 278.5±50 kcal.”
Confidence interval is an unknown function of the distribution, statistic, and sample size.

CI for $\bar{x} = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
The “basic bootstrap” (Efron, 1979)
The "basic bootstrap" (Efron, 1979)

| 2   |
| 318 |
| 214 |
| 318 |
| 2   |
| 326 |

Median: 266

278.5
The "basic bootstrap" (Efron, 1979)

<table>
<thead>
<tr>
<th>326</th>
<th>283</th>
<th>2</th>
<th>214</th>
<th>283</th>
<th>274</th>
</tr>
</thead>
</table>

The median of the list is 278.5.
The “basic bootstrap” (Efron, 1979)
Values of median corresponding to 2.5th and 97.5th percentiles => 95% confidence interval

The “basic bootstrap” (Efron, 1979)
“Donut contains $278.5\pm50$ kcal.”
If confidence interval not precise enough, how do we make it tighter?

Confidence interval = function of sample size…

For tighter intervals, ask more people!

"Donut contains 278.5±50 kcal."

VoxPL Algorithm

Increase sample size until either

1) interval meets user-defined “tightness” constraint

2) or budget exhausted

def numCalories(url: String) = Estimate (  
  confidenceInterval = SymmetricCI(50),  
  confidence = 0.95,  
  budget = 5.00,  
  text = "How many calories are in the food pictured?",  
  imageUrl = url,  
  statistic = L1Median  
)

(mandatory parameters are in red)
val cals = numCalories(breakfast) + numCalories(lunch)

Computing confidence intervals for composed functions: same procedure (because: bootstrap)

All functions expressible in VoxPL produce valid estimates & empirical confidence intervals

Also ensures that dynamic sample size calculation does not bias results (Bonferroni)
Calorie Counter

- 208 images of school lunches w/ground truth kcal (thanks to Joe Price @ Brigham Young U.)
- IM2Calories MAE: 152.95, SE: 15.61
- VoxPL MAE: 103.08, SE: 6.00
Calorie Counter

Fixed $n=200$, confidence interval ±41 kcal

Cost: $12 per plate!

VoxPL savings? Depends on constraint

- VoxPL dynamically finds right sample size - can be much cheaper!
  ±250, median cost: $0.32; ±50 cost: $1.28
AutoMan/VoxPL

Automatic budgeting, scheduling, and quality control

http://github.com/plasma-umass/AutoMan

DANIEL BAROWY
CHARLIE CURTSINGER
EMERY BERGER
ANDREW Mcgregor

UMass Amherst

DANIEL GOLDSTEIN
SIDDHARTH SURI
(VoxPL)

Microsoft
AutoMan debugger integrated with IntelliJ

Track progress
Identify bugs:
- unclear / too-difficult questions
- worker bias
- underpayment
Track overall progress
extra slides
Which one of these does not belong?

- apple
- celery
- cucumber
- orange
- banana
Is this a picture of a cowboy?
How many goats?