

Quality Control - part 2

Crowdsourcing and Human Computation

Instructor: Chris Callison-Burch

Website: crowdsourcing-class.org

Different Mechanisms for Quality Control

- Aggregation and redundancy
- Embedded gold standard data
- Economic incentives
- **Reputation systems**
- Statistical models

Reputation systems

- Mechanical Turk uses a reputation system
- Each Turker has a small number of variables associated with them, that are exposed to Requesters
- Past approval rate
- Number of HITs approved
- Has masters qualification (photo moderation/categorization master)

Pros and Cons of MTurk's reputation system

Pros	Cons
Gives a bit information about what other Requesters thought of a Worker	Reasons for rejections not shared; Weights all Requesters equally
Allows you to select Amazon's master's qualification, which is given to experienced Workers	It is not clear who gets the master's qual. No way to share other qualifications.
	Asymmetric: applies only to Workers, with no way to rate Requesters

Confederated Trust

- Acceptance rate doesn't show how good a worker is at a particular task
- Qualifications like the “photo moderation master’s” may show this
- However, there is no way to share this information with other requesters
- Lots of reinventing the wheel

Confederated Trust

- Do you think it would be useful to share qualifications among requesters?
- How would you do it?

Asymmetric reputation systems

- No way for Turkers to rate requesters, and see beforehand who is scrupulous
- Turkers have built their own external tools for this like TurkOpticon
- No way to see whether a Turkers high rating comes from good Requesters

Choose the best category for this government project (good english important)

Requester:

▼ [The Public Group](#)

HIT Expiration Date: Sep 10, 2013 (6 days)

Time Allotted: 60 minutes

communicativity:  1.17 / 5

generosity :  1.73 / 5

fairness :  1.39 / 5

promptness :  1.86 / 5

[What do these scores mean?](#)

Scores based on [81 reviews](#)

Terms of Service violation flags: 0

[Report your experience with this requester »](#)

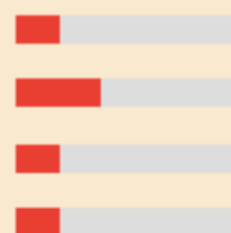
[Contact Us](#) | [Careers at Amazon](#) | [Developers](#) |

©2005-2013 Amazon.com, Inc. or its Affiliates



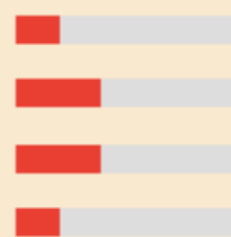
Took a leap of faith on this requester and was rewarded with a %50 reject rate and a broken search feature and no feedback. Would not recommend, even if you have thousands of HITs under your belt to cushion the inevitable rejections.

Aug 29 2013 | [KBH19](#) | [flag](#) | [comment](#)



Arbitrarily rejected over half of the hits I submitted, and then banned me from submitting any more hits for them. I suppose that's a blessing in disguise though, as I had no intention of doing any for them again after the first batch of rejections.

Aug 21 2013 | [bour...@g...](#) | [flag](#) | [comment](#)



Their HIT is very unclear. There is an option to browse for the result, but it does not work.

Aug 20 2013 | [jeff...@g...](#) | [flag](#) | [comment](#)

qualitative v quantitative

TurkOpticon's qualitative attributes	CrowdWorker's quantitative equivalents
promptness: How promptly has this requester approved your work and paid?	Expected time to payment: On average, how much time elapses between submitting work to this Requester and receiving payment?
generosity: How well has this requester paid for the amount of time their HITs take?	Average hourly rate: What is the average hourly rate that other Turker make when they do this requester's HITs?
fairness: How fair has this requester been in approving or rejecting your work?	Approval/rejection rates: What percent of assignments does this Requester approve? What percent of first-time Workers get any work rejected?
communicativity: How responsive has this requester been to communications or concerns you have raised?	Reasons for rejection: Archive of all of the reasons for Workers being rejected or blocked by this Requester.

Amazon's other reputation system

- Amazon has another reputation system in place for its online stores
- Amazon allows anyone to list and sell items through its site, and to set their own prices
- These can be individuals selling used goods, or independent 3rd party sellers who use Amazon to reach a larger customer base
- How does Amazon ensure good customer experience?

Feedback from buyers










- How satisfied were you with how your order was packaged and shipped?
- If you contacted the third-party seller, did you get good customer service and prompt resolution?
- Would you buy from this third-party seller again?



Westinghouse Lighting 7214100
Harmony Two-Light 48-Inch Two-Blade
Indoor Ceiling Fan, Brushed Nickel with
Opal Frosted Glass

by [Westinghouse](#) 

  ([42 customer reviews](#)) | [11 answered questions](#)

Price + Shipping	Condition	Seller Information	Buying Options
\$128.69 	New	amazon.com. In Stock. <ul style="list-style-type: none"> Free Two-day Shipping: Get it Wednesday, October 2 (order within) Domestic shipping rates and return policy.	 Add to cart or Turn on 1-Click to use your Amazon Prime benefits.
\$128.69 + \$24.32 shipping	New	 ★★★★★ 91% positive over the past 12 months. (12,817 total ratings) Ships in 1-2 business days. Expedited shipping available. Domestic shipping rates and return policy.	 Add to cart or Sign in to turn on 1-Click ordering.
\$128.69 + \$24.32 shipping	New	 ★★★★★ 90% positive over the past 12 months. (1,855 total ratings) Ships in 1-2 business days. Domestic shipping rates and return policy.	 Add to cart or Sign in to turn on 1-Click ordering.
\$148.99 + \$24.19 shipping	New	 ★★★★★ 97% positive over the past 12 months. (163,508 total ratings) Usually ships within 3 - 4 business days. Domestic shipping rates and return policy.	 Add to cart or Sign in to turn on 1-Click ordering.
\$202.90 FREE Shipping	New	DEL MAR <i>Fans & Lighting</i> ★★★★★ 98% positive over the past 12 months. (7,007 total ratings) Usually ships within 4 - 5 business days. Domestic shipping rates and return policy.	 Add to cart or Sign in to turn on 1-Click ordering.

Recent Feedback: ★★★★★

4.6 stars over the past 12 months (573 ratings)

[Previous Page](#) | [Next Page](#)

5/5: "good transaction, love the lock"
Sophia D., September 22, 2013

5/5: "Awesome experience "
Yadira Morejon, September 22, 2013

5/5: "Exactly what I needed, especially the color matched perfectly. Thank you."
mufasa, September 20, 2013

5/5: "Item was as described"
Thomas F., September 20, 2013

5/5: "Great seller, great item! Fast service too!!"
DB, September 20, 2013

2/5: "arrived bent"
ATD, September 20, 2013

Seller Response: We were not aware of any issue involving this customer's order. We have reached out to the customer to see how we may assist them in a return for a full refund, or a replacement of the damage item.

Date: September 23, 2013

5/5: "Shipping was a little pricey "

5/5: "Good price, high quality."
Joanna wang, September 18, 2013

5/5: "item was as descried seller prompt whith sevirce"
Thomas Howell, September 16, 2013

5/5: "works great "
Spencer , September 16, 2013

5/5: "just as described"
Theresa M., September 15, 2013

5/5: "Item was as described, value priced and works great."
Amanda S., September 14, 2013

1/5: "When a seller charges \$29.95 in shipping for a package weighing .2 lbs, they are gouging. The sponges cost less than \$18. I assume their profit is from the shipping. I could get this shipped for less than \$6.00! Never again"
Lana L Miller, September 14, 2013

Seller Response: We apologize the customer is not satisfied with the shipping charges. We have reached out to the customer and offered a discount to them as a one time courtesy.

Date: September 17, 2013

What are the economic implications of poor feedback?

\$128.69

+ \$24.32 shipping



★★★★★ **91% positive** over the past 12 months. (12,817 total ratings)

Ships in 1-2 business days. Expedited shipping available.
[Domestic shipping rates](#) and [return policy](#).

\$128.69

+ \$24.32 shipping



★★★★★ **90% positive** over the past 12 months. (1,855 total ratings)

Ships in 1-2 business days.
[Domestic shipping rates](#) and [return policy](#).

\$148.99

+ \$24.19 shipping



★★★★★ **97% positive** over the past 12 months. (163,508 total ratings)

Usually ships within 3 - 4 business days.
[Domestic shipping rates](#) and [return policy](#).

\$202.90

FREE Shipping



★★★★★ **98% positive** over the past 12 months. (7,007 total ratings)

Price premium

- Multiple sellers all selling the same item, but at different prices
- Price premium is the difference between a cheaper listing and a more expensive listing
- When someone opts for the more expensive item, even though it is identical, what is the reason for paying the premium?

Data-driven analysis

- Panos Ipeirotis harvested data from Amazon's website
- Gathered **transaction data** by repeatedly visiting listings (every 8 hours) and tracking when one item sold
- Gathered **reputation data** for each merchant. Complete history of numerical scores and text-based feedback

Data-driven analysis

- Data set gathered over half a year period
- Transaction data contains 1,078 merchants, 9,484 unique transactions and 107,922 price premiums
- Reputation data contains an average of 4,932 postings for each merchant

NLP + Economics

- Quantify the economics impact of sentiment of the feedback evaluations
- Using natural language processing techniques to derive semantic orientation and strength of comments

Method

- Each merchant's reputation is represented using a vector of n-dimensions $X = (X_1, X_2, \dots X_N)$
- Dimensions were 150 nouns and verbs, values of dimensions could be one of 140 modifiers
- X_1 is "delivery," X_2 is "packaging," X_3 is "service."
- Feedback 1 "I was impressed by the speedy delivery! Great service!": (speedy ; NULL; great)
- Feedback 2 "The item arrived in awful packaging, and the delivery was slow": (slow ; awful ; NULL)

Method

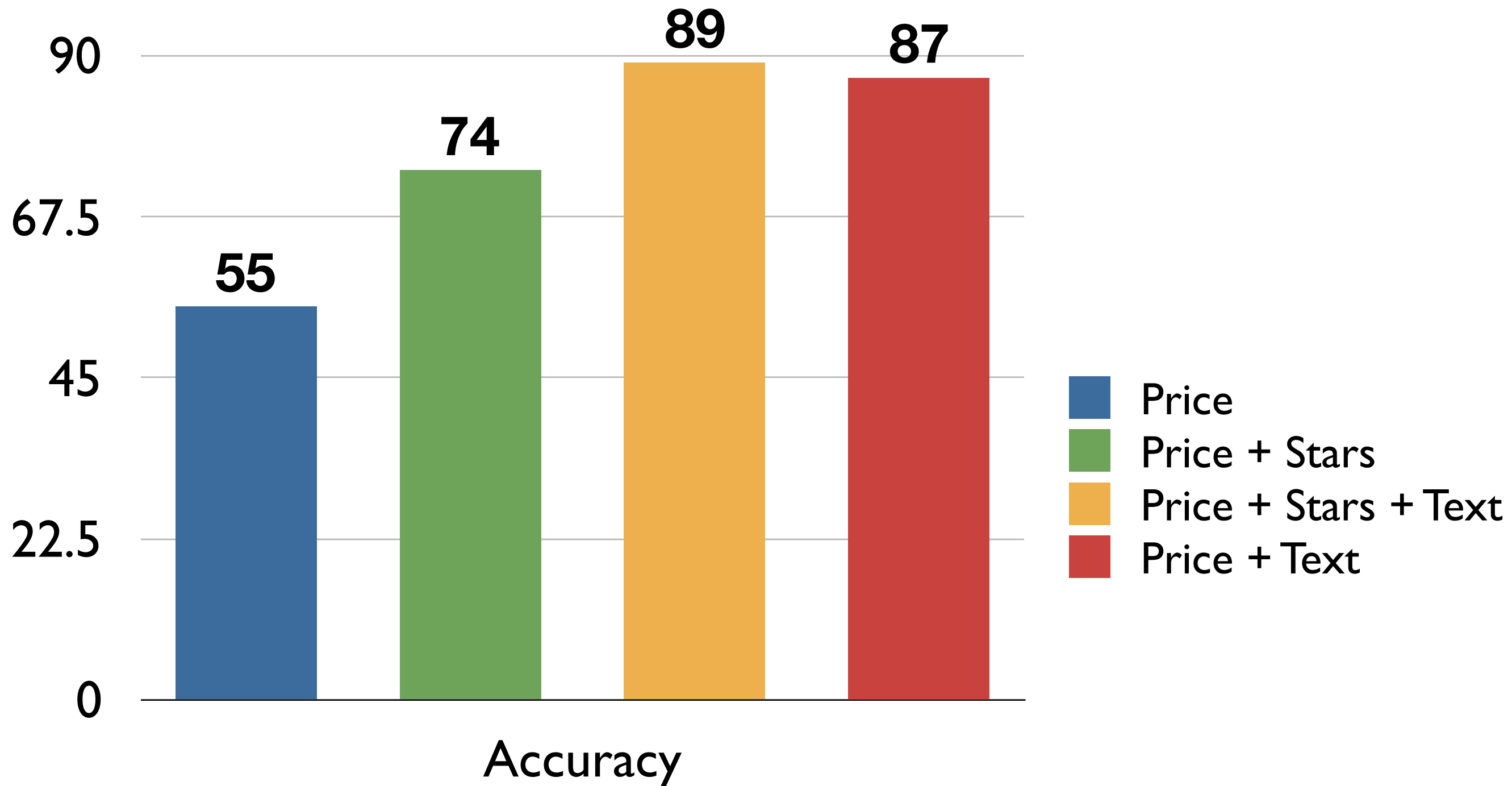
- Construct a matrix out of all of the feedback for a seller
- Weight the more recent feedback more heavily
- Calculate how the values of each dimension effect the price premium
- Use least-squares regression with fixed effects to predict the price premium

Highest scoring phrases

wonderful experience	\$5.86
outstanding seller	\$5.76
excellant service	\$5.27
lightning delivery	\$4.84
highly recommended	\$4.15
best seller	\$3.80
perfectly packaged	\$3.74
excellent condition	\$3.53
excellent purchase	\$3.22
excellent seller	\$2.70
excellent communication	\$2.38
perfect item	\$1.92
terrific condition	\$1.87
top quality	\$1.67
awesome service	\$1.05
A+++ seller	\$1.03
great merchant	\$0.93

never received	-\$7.56
defective product	-\$6.82
horrible experience	-\$6.79
never sent	-\$6.69
never recieved	-\$5.29
bad experience	-\$5.26
cancelled order	-\$5.01
never responded	-\$4.87
wrong product	-\$4.39
not as advertised	-\$3.93
poor packaging	-\$2.92
late shipping	-\$2.89
wrong item	-\$2.50
not yet received	-\$2.35
still waiting	-\$2.25
wrong address	-\$1.54
never buy	-\$1.48

Predicting the merchant who makes the sale



Challenges for Reputation Systems

- Not enough people participate
- Feedback tends to be overwhelmingly positive
- Reports can be dishonest
- Reputation systems are undermined if people can change identities easily
- People can milk a good reputation

Insufficient participation

- Giving feedback for a reputation system contributes to the public good
- However, after some information is available it is easy for people to be "free riders" without contributing anything
- Early raters take on a transaction cost (Yelpers risk going to bad restaurants with no reviews)
- Solutions?

Overwhelmingly positive feedback

- 99% of all feedback on eBay is positive
- Part of the problem is **reciprocity**
- Sellers and buyers evaluate each other
- Positive ratings are given in the hopes of getting positive ratings in return
- Negative ratings are avoided for fear of getting negative feedback as retaliation

Dishonest reports

- **Ballot stuffing** - a seller colludes with buyers to give unfairly high ratings
- **Bad mouthing** - collusion to give negative feedback about competitors that they want to drive out of the market

Identity changes

- **Cheap pseudonyms** - easy to disappear and re-register under a new identity with almost zero cost
- Can misbehave without paying consequences toward reputation

Value imbalance exploitations

- People who want to commit fraud could first invest in building a good reputation
- Ebay exploit: "Riddle for 1¢. No shipping. Positive feedback"
- Sellers would take a 29¢ loss to build up positive reputation quickly

Challenges for Crowdsourcing Markets

- Reciprocal systems are worse than 1-sided systems in e-commerce.
- Only the sellers are likely to behave opportunistically. No need for reciprocal evaluation.
- In crowdsourcing, both sides can be fraudulent. So reciprocal markets are important, but they are hard to get right!

Challenges for Crowdsourcing Markets

- In e-commerce markets, it is straightforward for buyers to evaluate the quality of the product when they receive it.
- In crowdsourcing markets, verifying the correct answer is sometimes as costly as producing it.
- This has the potential to significantly reduce participation and/or accuracy of reviews

Challenges for Crowdsourcing Markets

- No "price premium" for high quality workers
- In e-commerce markets, sellers with a good reputation can sell their goods at a relatively high price (premium)
- In crowdsourcing, the requester sets the price, and this is typically the same for all workers

Different Mechanisms for Quality Control

- Aggregation and redundancy
- Embedded gold standard data
- Economic incentives
- Reputation systems
- **Statistical models**

Expectation Maximization algorithm

- EM is an algorithm for finding the probabilities of unobserved variables
- We will use it to estimate how accurate workers' labels are, and infer how good each worker is
- This is more sophisticated than voting

Dawid and Skene (1977)

- *Maximum Likelihood Estimation of Observer Error-rates using the EM Algorithm*
- Examined application to medical diagnosis
- Patients are sometimes treated by multiple physicians, who can give different diagnoses
- Why? Doctors may have different questions.
Patient may describe history differently.
Doctors may classify symptoms differently

Observer Error

- Given that different doctors have different opinions, they can't all be right.
- How often do individual physicians suffer from “observer error”? Are their errors systematic?
- Answers depend on the “true” diagnosis

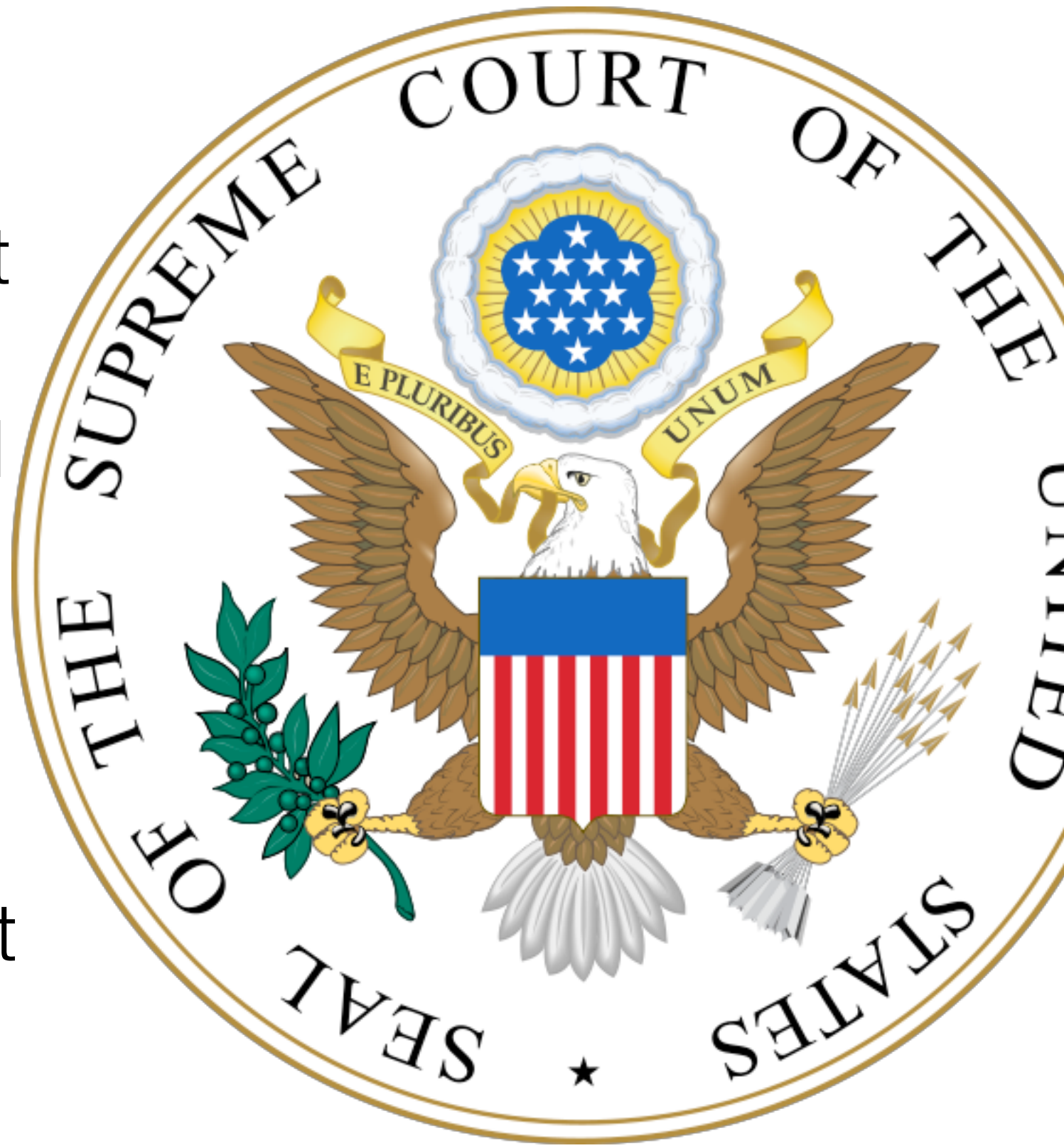
Observer Error

- Observer error would be easy to calculate if we had ground truth
- Simply count the misdiagnoses and divide by the total number of diagnoses
- However, sometimes it is impossible to know what diagnosis is correct. Same set of symptoms can arise from multiple root causes.

“I know it when I see it”

I shall not today attempt further to define "hard-core pornography"; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it.

—Justice Potter Stewart



url	worker1	worker2	worker3	worker4	worker5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

Solution?

- Can't have Justice Stewart rule on everything
- Instead, we will apply Dawid and Skene's EM algorithm, which iteratively
 1. Estimates the correct answers, using labels from multiple workers, and accounts for the quality of each worker
 2. Estimates the quality of the workers by comparing the submitted answers to the inferred correct answers

Inputs

- a set of **N** objects $o_1 \dots o_N$
sunnyfun.com, sex-mission.com,
google.com, youporn.com, yahoo.com
- a set of **L** possible labels:
{porn, not porn}
- Labels for each object by **K** workers
worker1, worker2, worker3, worker4,
worker5

Goal 1

- Recover the true class label $\mathbf{T}(O_n)$ for each object O_n when “gold” truth is unknown
- Since the true labels are not known / never directly observed, they are called ***latent*** variables

Goal 2

- For each worker who contributed labels, calculate their accuracy or reliability
- To calculate accuracy show how often they mistakenly choose one label when a different one is the actual truth

Chicken and egg problem

- If we knew what the **true class labels** were for each object for each object, then we could compute each Turker's accuracy
- If we had **accuracies for every Turker**, then we could infer what the true label for each object should be

Input: Labels $l[k][n]$ from worker (k) to object o_n ,

Output: Confusion matrix $\pi_{ij}^{(k)}$ for each worker (k), Correct labels $T(o_n)$ for each object o_n , Class priors $Pr\{C\}$ for each class C

- 1 Initialize error rates $\pi_{ij}^{(k)}$ for each worker (k) (e.g., assume each worker is perfect);
- 2 Initialize correct label for each object $T(o_n)$ (e.g., using majority vote);
- 3 **while** *not converged* **do**
- 4 Estimate the correct label $T(o_n)$ for each object, using the labels $l[\cdot][n]$ assigned to o_n by workers, weighting the votes using the error rates $\pi_{ij}^{(k)}$;
- 5 Estimate the error rates $\pi_{ij}^{(k)}$, for each worker (k), using the correct labels $T(o_n)$ and the assigned labels $l[k][n]$;
- 6 Estimate the class priors $Pr\{C\}$, for each class C ;
- 7 **end**
- 8 **return** *Estimated error rates $\pi_{ij}^{(k)}$, Estimated correct labels $T(o_n)$, Estimated class priors $Pr\{C\}$*

Algorithm 1: The EM algorithm for worker quality estimation.

Input

url	worker1	worker2	worker3	worker4	worker5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

Output: “True” Labels

url	True Labels
sunnyfun.com	not
sex-mission.com	porn
google.com	not
youporn.com	porn
yahoo.com	not

Output: Worker Accuracies

		Guess	
Truth	w1	porn	not
	porn	100%	0%
	not	67%	33%

		Guess	
Truth	w2	porn	not
	porn	100%	0%
	not	33%	67%

		Guess	
Truth	w3	porn	not
	porn	100%	0%
	not	0%	100%

		Guess	
Truth	w4	porn	not
	porn	100%	0%
	not	0%	100%

		Guess	
Truth	w5	porn	not
	porn	50%	50%
	not	100%	0%

Initialize

Guess

Truth	w1	porn	not
	porn	100%	0%
	not	0%	100%

	w2	porn	not
	porn	100%	0%
	not	0%	100%

	w3	porn	not
	porn	100%	0%
	not	0%	100%

Guess

Truth	w4	porn	not
	porn	100%	0%
	not	0%	100%

	w5	porn	not
	porn	100%	0%
	not	0%	100%

Initialize with majority vote

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex- mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
40%	60%
100%	0%
40%	60%
80%	20%
40%	60%

Pretend your labels are correct

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
0%	100%
100%	0%
0%	100%
100%	0%
0%	100%

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	0	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	1	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	1	0
not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	2	0
not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	2	0
not	2	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w1	
	porn	not
porn	100%	0%
not	67%	33%

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
0	1
1	0
0	1
1	0
0	1

Re-Calculate
Worker
Scores

		Guess	
Truth	w2	porn	not
	porn	0	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

		Guess	
Truth	w2	porn	not
	porn	1	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth	Guess	
	w2	
	porn	not
porn	1	0
not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

		Guess	
Truth	w2	porn	not
	porn	2	0
	not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth	Guess	
	w2	
	porn	not
porn	2	0
not	1	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth	Guess	
	w2	
	porn	not
porn	100%	0%
not	33%	67%

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
0	1
1	0
0	1
1	0
0	1

Re-Calculate
Worker
Scores

Truth

Guess

w3	porn	not
porn	0	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth

Guess

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth

Guess

w3	porn	not
porn	1	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth

Guess

w3	porn	not
porn	2	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth

Guess

w3	porn	not
porn	2	0
not	0	3

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth

Guess

w3	porn	not
porn	100%	0%
not	0%	100%

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
0	1
1	0
0	1
1	0
0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w4	
	porn	not
porn	0	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w4	
	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Guess

Truth

w4	porn	not
porn	1	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Guess

Truth

w4	porn	not
porn	2	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Guess

Truth

w4	porn	not
porn	2	0
not	0	3

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w4	
	porn	not
porn	100%	0%
not	0%	100%

url	worker 1	worker 2	worker 3	worker 4	worker 5
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

porn	not porn
0	1
1	0
0	1
1	0
0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w5	
	porn	not
porn	0	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

	Guess	
	w5	
	porn	not
Truth	porn	10
	not	10

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w5	
	porn	not
porn	1	0
not	2	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

		Guess	
		w5	
Truth	porn	1	1
	not	2	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate
Worker
Scores

Truth	Guess	
	w5	
	porn	not
porn	1	1
not	3	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

Re-Calculate Worker Scores

Truth	Guess	
	w5	
	porn	not
porn	1	1
not	3	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex-mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

Re-Calculate Worker Scores

Truth	Guess	
	w5	
	porn	not
porn	50%	50%
not	100%	0%

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex-mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

Update Estimates for True Labels

	Guess	
	w1	
	porn	not
Truth	porn	1.00 0.00
	not	0.67 0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	0.67
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

	Guess	
	w2	
	porn	not
Truth	porn	1.00 0.00
	not	0.33 0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	1.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Guess

Update
Estimates for
True Labels

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	2.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Guess

Update
Estimates for
True Labels

Truth

w4	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	3.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

		Guess	
		w5	
Truth	porn	0.5	0.5
	not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.50	4.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.50	4.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w1	
	porn	not
porn	1.00	0.00
not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		1.00	0.67
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

		Guess	
Truth	w2	porn	not
	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		2.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Guess

Update
Estimates for
True Labels

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		3.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w4	
	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		4.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

		Guess	
		w5	
Truth	porn	0.5	0.5
	not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		4.50	2.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		4.50	2.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

		Guess	
Truth	w1	porn	not
	porn	1.00	0.00
	not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		0.00	0.33
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

	Guess	
	w2	
	porn	not
Truth	porn	1.00 0.00
	not	0.33 0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	0.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Guess

Update
Estimates for
True Labels

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	1.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w4	
	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	2.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w5	
	porn	not
porn	0.5	0.5
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.50	3.76
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.50	3.76
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w1	
	porn	not
porn	1.00	0.00
not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		1.00	0.67
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

	Guess	
	w2	
	porn	not
Truth	porn	1.00 0.00
	not	0.33 0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		2.00	1.00
yahoo.com	porn	not	not	not	porn			

Guess

Update
Estimates for
True Labels

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		3.00	1.00
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w4	
	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.00	1.00
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w5	
	porn	not
porn	0.5	0.5
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.50	1.00
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.50	1.00
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn			

Update Estimates for True Labels

Truth	Guess	
	w1	
	porn	not
porn	1.00	0.00
not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	0.67

Update Estimates for True Labels

		Guess	
Truth	w2	porn	not
	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	1.34

Guess

Update
Estimates for
True Labels

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	2.34

Update Estimates for True Labels

Truth	Guess	
	w4	
	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	3.34

Update Estimates for True Labels

Truth	Guess	
	w5	
	porn	not
porn	0.5	0.5
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.50	4.34

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5			
						porn	not porn	
sunnyfun.com	porn	not	not	not	porn	26%	74%	
sex- mission.com	porn	porn	porn	porn	porn	69%	31%	
google.com	not	porn	not	not	porn	29%	71%	
youporn.com	porn	porn	porn	porn	not	82%	18%	
yahoo.com	porn	not	not	not	porn		1.50	4.34

Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5			
						porn	not porn	
sunnyfun.com	porn	not	not	not	porn	26%	74%	
sex- mission.com	porn	porn	porn	porn	porn	69%	31%	
google.com	not	porn	not	not	porn	29%	71%	
youporn.com	porn	porn	porn	porn	not	82%	18%	
yahoo.com	porn	not	not	not	porn		26%	74%

Iteration 0

url	porn	not porn
sunnyfun.com	40%	60%
sex-mission.com	100%	0%
google.com	40%	60%
youporn.com	80%	20%
yahoo.com	40%	60%

Guess

Truth	w1	porn	not
	porn	100%	0%
	not	0%	100%
	w2	porn	not
	porn	100%	0%
	not	0%	100%
	w3&4	porn	not
	porn	100%	0%
	not	0%	100%
	w5	porn	not
	porn	100%	0%
	not	0%	100%

Iteration 1

url	porn	not porn
sunnyfun.com	26%	74%
sex- mission.com	69%	31%
google.com	29%	71%
youporn.com	82%	18%
yahoo.com	26%	74%

		Guess	
Truth	w1	porn	not
	porn	100%	0%
	not	67%	33%
	w2	porn	not
	porn	100%	0%
	not	33%	67%
	w3&4	porn	not
	porn	100%	0%
	not	0%	100%
	w5	porn	not
	porn	50%	50%
	not	100%	0%

Repeat until convergence

- You can continue to iterate until your values converge
- For this example, we converge after the first iteration

Question

- How would you use gold standard data in the EM process?

EM Algorithm

- Re-Calculate Worker Scores over two steps:
 1. Estimate the probability that each answer is correct, using labels from multiple workers weighted by the probability that they are correct
 2. Estimate the quality of the workers by comparing their submitted answers to the inferred correct answers

Confusion Matrix gives us worker error

- From the confusion matrix we can measure the overall error rate for each worker
- Sum of the non-diagonal elements of the confusion matrix (weighted by the priors)
- This results in a single, scalar value as the quality score for each worker

Worker Error

w1	porn	not
porn	100%	0%
not	100%	0%

100

w2	porn	not
porn	100%	0%
not	33%	67%

33

w3	porn	not
porn	100%	0%
not	0%	100%

0

w4	porn	not
porn	0%	100%
not	100%	0%

200

Is worker5 the worst?

url	worker1	worker2	worker3	worker4	worker5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

Advanced Topics

- Bias versus error
- How noisy can the workers be and still allow us to still converge to a correct solution?

Bias versus Error

- Error rate alone is not sufficient to measure the inherent value of a worker.
- For example, workers may be careful but biased
- In a non-binary case, this is more apparent
- What if instead of asking our workers to label sites porn or not porn, we asked them to label the G, PG, R, X?

Bias versus Error

- Parents with young children tend to be more conservative
- They tend to classify PG-rated sites as R-rated sites, and R-rated sites as X-rated.
- Such workers give consistently and predictably incorrect answers
- It is possible to automatically correct for bias

Implications

- Unlike with spammers, with biased workers it is possible to “reverse” the errors
- We can recover a label assignment of much higher quality
- In the presence of systematic bias, the naive measurement of error rate results in underestimates of the true quality of the worker
- This potentially leads to incorrect rejections and blocks of legitimate workers

For more details

- Check out two papers by Panos Ipeirotis and his collaborators
- *Managing Crowdsourcing Workers* discusses separating error and bias
- *Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers* discusses how noisy judgements can be, with us still getting good quality results