Lecture 6

Label Aggregation Wrap-Up & Biases in Human-Generated Data

Chien-Ju (CJ) Ho

Logistics: Presentation

#	Date	Topic	Presenters
#1	Sep 27	Incentive Design: Financial Incentives	Daisy Wang, Xinyi Ye
#2	Sep 29	Incentive Design: Badges and Attention	CJ
#3	Oct 4	Application: Darpa Network Challenge	Yiding Tao, Xiangyu Chen
#4	Oct 6	Application: Prediction Markets	Qihang Huang, Zheng Wang and Zhuomin Li
#5	Oct 18	Workflow Design	CJ
#6	Oct 20	Expert Crowdsourcing and Teams	Danielle Beaulieu, Kaushik Dutta
#7	Oct 25	Non-Independent Work and Argumentation	Cenhao Li, Ruiwei Xiao, Yang Yi
#8	Nov 3	Fairness in AI	David Sarpong, Alex Wollam
#9	Nov 8	Human Perceptions of Fairness	Aayush Dhakal, Subash Khanal
#10	Nov 10	Ethical Decision Making and Participatory Design	Tejas Mattur, Run Zhang, Jacob Dodd
#11	Nov 17	Interpretable Machine Learning	Ming Gao, Boyan Tian, Jiayi Zhang
#12	Nov 29	Human-AI Team (1)	Ruowen Xu, Yucen Zhong
#13	Dec 1	Human-AI Team (2)	Jiajun Sun, Xianchun Zeng, Miao Qin

Logistics: Presentation

- For presenters:
  - Give a **55~60 min** presentation based on the **required reading** and at least **two optional reading** (3 optional readings for 3-person groups) of a lecture
    - The papers are the “backbone” of the presentation
  - Prepare **2 reading questions** for the required reading
  - Prepare **at least 2 discussion sessions**
  - Lead the discussion for the discussion sessions
  - <u>Template format</u> (if you are not sure what to do):
    - Explain the required reading (5~10 min)
    - Discussion session (5~10 min)
    - Discussion on the optional readings (15~20 min)
    - Another discussion session (5~10 min)
    - Discussion on the optional readings (15~20 min)
    - A short summary (3~5 min)
    - Feel free to be creative and include materials outside of the papers

Logistics: Presentation

- For presenters:
  - Talk to me **one week before your presentation**
    - Default time: talk to me after class
  - You need to be ready for the following before meeting with me
    - Finish reading the papers
    - A structure of your presentation
    - Topics for the discussion sessions
    - Two reading questions for the required reading

Logistics: Presentation

- For non-presenters:
  - Read the required reading and submit reviews.
  - Attend the lecture and engage in discussion.
  - Fill in peer review forms (probably an online form)
    - Comments are not anonymous to me but will be anonymous to the presenters.
    - Anonymized comments will be given to the presenters.
    - Please give constructive comments to help each other. Presentation is a very helpful skill for your future career.

Logistics: Assignments and Project Proposal

- Assignments
  - Assignment 1 is due this Friday
  - Assignment 2 is posted and due Sep 30
- Project proposal
  - Due next Friday (Sep 23). No late submissions.
  - Requirements
    - Title / 1-to-2 paragraph descriptions / citing one paper
  - A list of example/past projects is posted on the course website
  - You are encouraged to start with a research project. You will have the opportunities to make it a literature survey later (before milestone 2).

Logistics: Project Proposal

- Application:
  - Prototyping a system that combine humans and computation to solve tasks
  - Start small, so you can showcase the results with a small number of users
  - Example: trip planning, nutrition analysis, …
- Designing mechanisms/systems with human involved
  - Assume certain human behavioral models, design systems/mechanisms that maximize the objective
  - Example:
    - Design reputation systems to encourage good behavior
    - Design news recommendation to mitigate polarization
  - Key: assume some user models
    - Conduct theoretical analysis, or
    - Run simulations that assume users behave as the model suggests
    - Could study multiple user models, and explore how that impacts the design

Logistics: Project Proposal

- Impacts of human behavior to standard systems
  - We have looked at label aggregation and will look at incentive design that assume standard human behavior
  - Explore what happens when humans don’t behave according to the assumption
  - Study the possible manipulation or adversarial attack to sabotage the system
  - Study the design of robust systems that are robust to attacks
- Understanding human behavior
  - Crawl data from the Web or utilize the public datasets
  - Study how humans behave using the data
  - You might also run behavioral experiments itself
    - Not recommended (due to logistical complexity), and please talk to me early if you want to do so

Logistics: Project Proposal

- There is flexibility on the project topic
  - Need to be relevant to the course and have strong human components
  - I’ll make the final call on whether it’s related to the course
- You can still change the topic before milestone 1
- You can convert a research project to an extensive literature survey before milestone 2

Lecture Today

What We Learned So Far in Label Aggregation

- EM-based methods (Mainstream methods)
  - Empirically performs well
  - Relatively computationally efficient
  - No theoretical guarantee
- Matrix-based methods (A taste on theory-grounded work)
  - Computationally more expensive
  - Comes with theoretical guarantee
  - Require some “potentially unreasonable” assumptions for the analysis
- There are some other approaches

One more example:

Learning from the Wisdom of Crowd by Minimax Entropy. Zhou et al. NIPS 2012.

Entropy (Information Entropy)

- Consider a random variable X with n possible values
- The probability for each value i happening is P_i

- Information entropy (Shannon entropy)

$$H(X) = -\sum_{i=1}^{n} P_i \log P_i$$

What are the interpretations of entropy?

Higher entropy => More uncertainty => Higher unpredictability

Principle of Maximum Entropy

"the probability distribution which best represents the current state of knowledge is the one with largest entropy"

- Consider a dice with 6 faces
  - Without any knowledge, what's your best bet on the probability of 1~6 happening
  - Assume you are told the probability of 3 happening is ½, what's your best bet on the probability of the rest numbers happening?

How does this apply to label aggregation?

- We are trying to infer
  - true task labels
  - worker skills
  - and maybe other parameters

- Principle of Maximum Entropy
  - Worker skills are often modeled as “probability distributions”
  - Given observed labels, we can infer worker skills that “maximize entropy”
  - We can then infer labels that minimizes uncertainty

Setting

Goal: Given $\vec{z}$, how to infer $\vec{\pi}$ and $\vec{y}$?

Observations

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{z}_{1,1}$	$\vec{z}_{1,2}$	$\vec{z}_{1,3}$	…	$\vec{z}_{1,n}$
Worker 2	$\vec{z}_{2,1}$	$\vec{z}_{2,2}$	$\vec{z}_{2,3}$	…	$\vec{z}_{2,n}$
Worker 3	$\vec{z}_{3,1}$	$\vec{z}_{3,2}$	$\vec{z}_{3,3}$	…	$\vec{z}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{z}_{m,1}$	$\vec{z}_{m,2}$	$\vec{z}_{m,3}$	…	$\vec{z}_{m,n}$

Underlying distribution

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{\pi}_{1,1}$	$\vec{\pi}_{1,2}$	$\vec{\pi}_{1,3}$	…	$\vec{\pi}_{1,n}$
Worker 2	$\vec{\pi}_{2,1}$	$\vec{\pi}_{2,2}$	$\vec{\pi}_{2,3}$	…	$\vec{\pi}_{2,n}$
Worker 3	$\vec{\pi}_{3,1}$	$\vec{\pi}_{3,2}$	$\vec{\pi}_{3,3}$	…	$\vec{\pi}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{\pi}_{m,1}$	$\vec{\pi}_{m,2}$	$\vec{\pi}_{m,3}$	…	$\vec{\pi}_{m,n}$

- Components
  - Workers $i = 1, \dots, m$
  - Tasks $j = 1, \dots, n$
  - Labels $k = 1, \dots, c$
- Worker labels $\vec{z}_{i,j} = (z_{i,j,1}, \dots, z_{i,j,c})$
  - $z_{i,j,k} = 1$ if worker i label task j as class k
  - $z_{i,j,k} = 0$ otherwise
- True labels $\vec{y}_j = (y_{i,1}, \dots, y_{j,c})$
  - $y_{j,l} = 1$ if task j's label is l
  - $y_{j,l} = 0$ otherwise
- Worker skills: $\vec{\pi}_{i,j} = (\pi_{i,j,1}, \dots, \pi_{i,j,c})$
  - $z_{i,j,k}$: probability for worker i label task j as class k

Apply the Maximum Entropy Principle

- Assume true labels $\vec{y}_j$ are given, how to infer worker skills $\vec{\pi}$?
- Choose $\vec{\pi}$ that maximizes entropy subject to the observations of $\vec{z}$

- Choose $\vec{\pi}$ that maximizes entropy subject to the observations of $\vec{z}$

$$\max_{\pi} \boxed{-\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{c}\pi_{ijk}\ln\pi_{ijk}}$$

Entropy

s.t.

$$\boxed{\sum_{k=1}^{c}\pi_{ijk}=1,\ \forall i,j,\ \pi_{ijk}\geq 0,\ \forall i,j,k.}$$

Probability constraints

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{z}_{1,1}$	$\vec{z}_{1,2}$	$\vec{z}_{1,3}$	…	$\vec{z}_{1,n}$
Worker 2	$\vec{z}_{2,1}$	$\vec{z}_{2,2}$	$\vec{z}_{2,3}$	…	$\vec{z}_{2,n}$
Worker 3	$\vec{z}_{3,1}$	$\vec{z}_{3,2}$	$\vec{z}_{3,3}$	…	$\vec{z}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{z}_{m,1}$	$\vec{z}_{m,2}$	$\vec{z}_{m,3}$	…	$\vec{z}_{m,n}$

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{\pi}_{1,1}$	$\vec{\pi}_{1,2}$	$\vec{\pi}_{1,3}$	…	$\vec{\pi}_{1,n}$
Worker 2	$\vec{\pi}_{2,1}$	$\vec{\pi}_{2,2}$	$\vec{\pi}_{2,3}$	…	$\vec{\pi}_{2,n}$
Worker 3	$\vec{\pi}_{3,1}$	$\vec{\pi}_{3,2}$	$\vec{\pi}_{3,3}$	…	$\vec{\pi}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{\pi}_{m,1}$	$\vec{\pi}_{m,2}$	$\vec{\pi}_{m,3}$	…	$\vec{\pi}_{m,n}$

- Choose $\vec{\pi}$ that maximizes entropy subject to the observations of $\vec{z}$

Consistency constraints

$$\max_{\pi} \quad -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{c} \pi_{ijk} \ln \pi_{ijk}$$

$$\text{s.t.} \quad \boxed{\sum_{i=1}^{m} \pi_{ijk} = \sum_{i=1}^{m} z_{ijk}, \ \forall j,k,} \ \boxed{\sum_{j=1}^{n} y_{jl}\pi_{ijk} = \sum_{j=1}^{n} y_{jl} z_{ijk}, \ \forall i,k,l,}$$

$$\sum_{k=1}^{c} \pi_{ijk} = 1, \ \forall i,j, \ \pi_{ijk} \geq 0, \ \forall i,j,k.$$

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{z}_{1,1}$	$\vec{z}_{1,2}$	$\vec{z}_{1,3}$	…	$\vec{z}_{1,n}$
Worker 2	$\vec{z}_{2,1}$	$\vec{z}_{2,2}$	$\vec{z}_{2,3}$	…	$\vec{z}_{2,n}$
Worker 3	$\vec{z}_{3,1}$	$\vec{z}_{3,2}$	$\vec{z}_{3,3}$	…	$\vec{z}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{z}_{m,1}$	$\vec{z}_{m,2}$	$\vec{z}_{m,3}$	…	$\vec{z}_{m,n}$

	Task 1	Task 2	Task 3	…	Task n
Worker 1	$\vec{\pi}_{1,1}$	$\vec{\pi}_{1,2}$	$\vec{\pi}_{1,3}$	…	$\vec{\pi}_{1,n}$
Worker 2	$\vec{\pi}_{2,1}$	$\vec{\pi}_{2,2}$	$\vec{\pi}_{2,3}$	…	$\vec{\pi}_{2,n}$
Worker 3	$\vec{\pi}_{3,1}$	$\vec{\pi}_{3,2}$	$\vec{\pi}_{3,3}$	…	$\vec{\pi}_{3,n}$
…	…	…	…	…	…
Worker m	$\vec{\pi}_{m,1}$	$\vec{\pi}_{m,2}$	$\vec{\pi}_{m,3}$	…	$\vec{\pi}_{m,n}$

Solving the Optimization

- Given true labels y, we use maximum entropy to find π

=> For every set of true labels y, we obtain π and the corresponding entropy

- How to decide the true labels y?
  - Higher entropy => higher uncertainty
  - Choosing labels that minimize uncertainty/entropy

- Minimax entropy

$$\min_{y} \max_{\pi} \quad -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{c} \pi_{ijk} \ln \pi_{ijk}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} \pi_{ijk} = \sum_{i=1}^{m} z_{ijk}, \ \forall j, k, \ \sum_{j=1}^{n} y_{jl}\pi_{ijk} = \sum_{j=1}^{n} y_{jl} z_{ijk}, \ \forall i, k, l,$$

$$\sum_{k=1}^{c} \pi_{ijk} = 1, \ \forall i, j, \ \pi_{ijk} \geq 0, \ \forall i, j, k, \ \sum_{l=1}^{c} y_{jl} = 1, \ \forall j, \ y_{jl} \geq 0, \ \forall j, l.$$

An interesting way of looking at label aggregation

- Finding the labels/distribution with minimax entropy

- Can we incorporate models of label generation?
  - e.g., Tasks are homogeneous
  - e.g., Tasks have different difficulty levels

- Express them as additional constraints

Additional Details on the Technical Insights

- Perform reasonably well in practice

Method	Dogs	Web
Minimax Entropy	84.63	88.05
Dawid & Skene	84.14	83.98
Majority Voting	82.09	73.07
Average Worker	70.60	37.05

- The dual formulation gives nice insights
  - One set of dual variables represent worker skills
  - Another set of dual variable represent task difficulties

A Recap on Label Aggregation

The Approaches We Covered

- EM-Based methods (The mainstream approach)
  - Develop models of label generation
  - Write down the likelihood function
  - Using EM algorithms to optimize likelihood
- Matrix-based method
  - Perform SVD, using the top left singular vector as the prediction
- Others
  - Minimax entropy
  - And more…

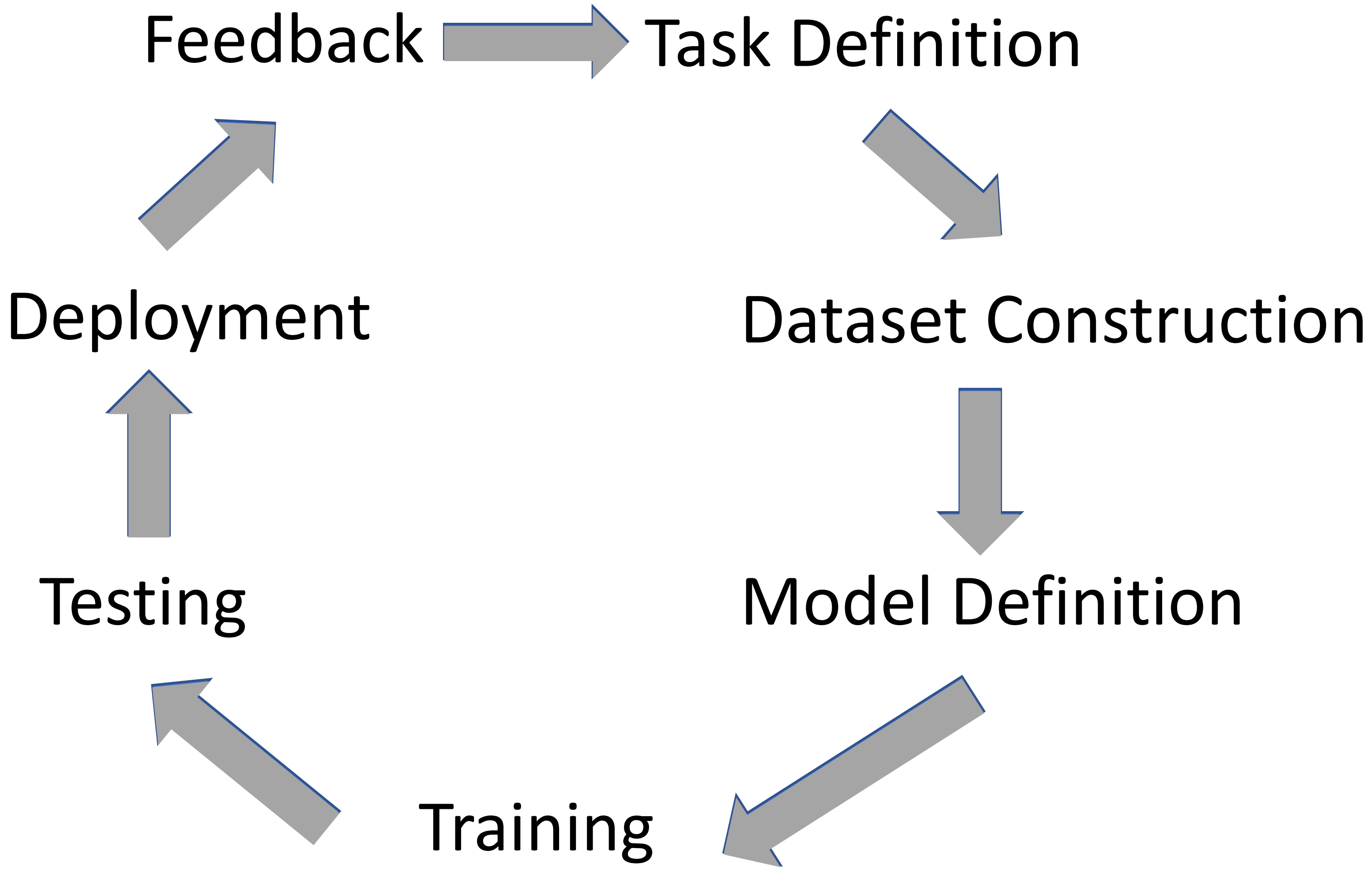
General Discussion on Label Aggregation

- Common assumption: each label is i.i.d. drawn from some distribution
- This assumption enables tons of papers applying statistics/learning techniques in crowdsourcing (low-hanging fruit)

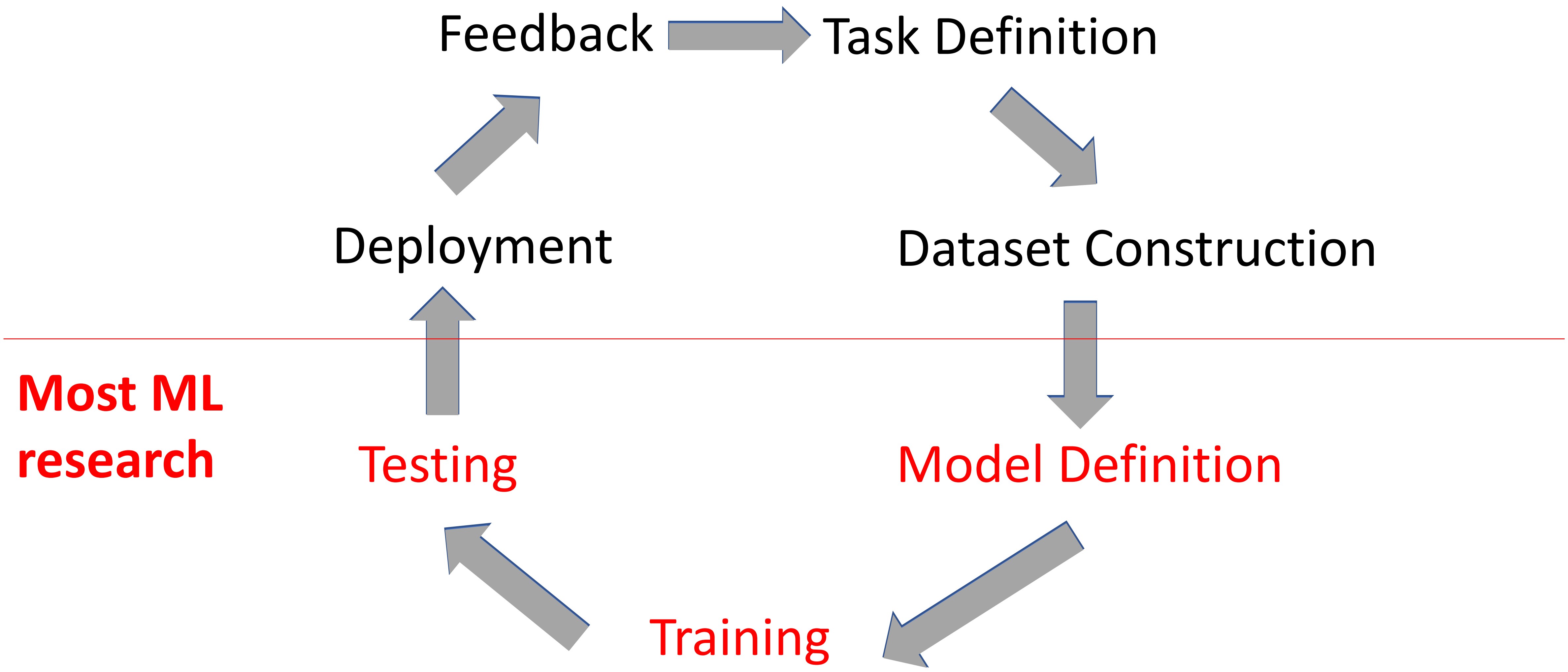
- Discussion
  - What other assumptions have been made in the papers you read?
  - Under what scenarios do you think this (and/or other assumptions) is reasonable?
  - Is there any assumption you think we should try to relax in this line of research.
  - If you need to keep working on label aggregation, what would you propose to do?

Concerns on
Human as Data Sources

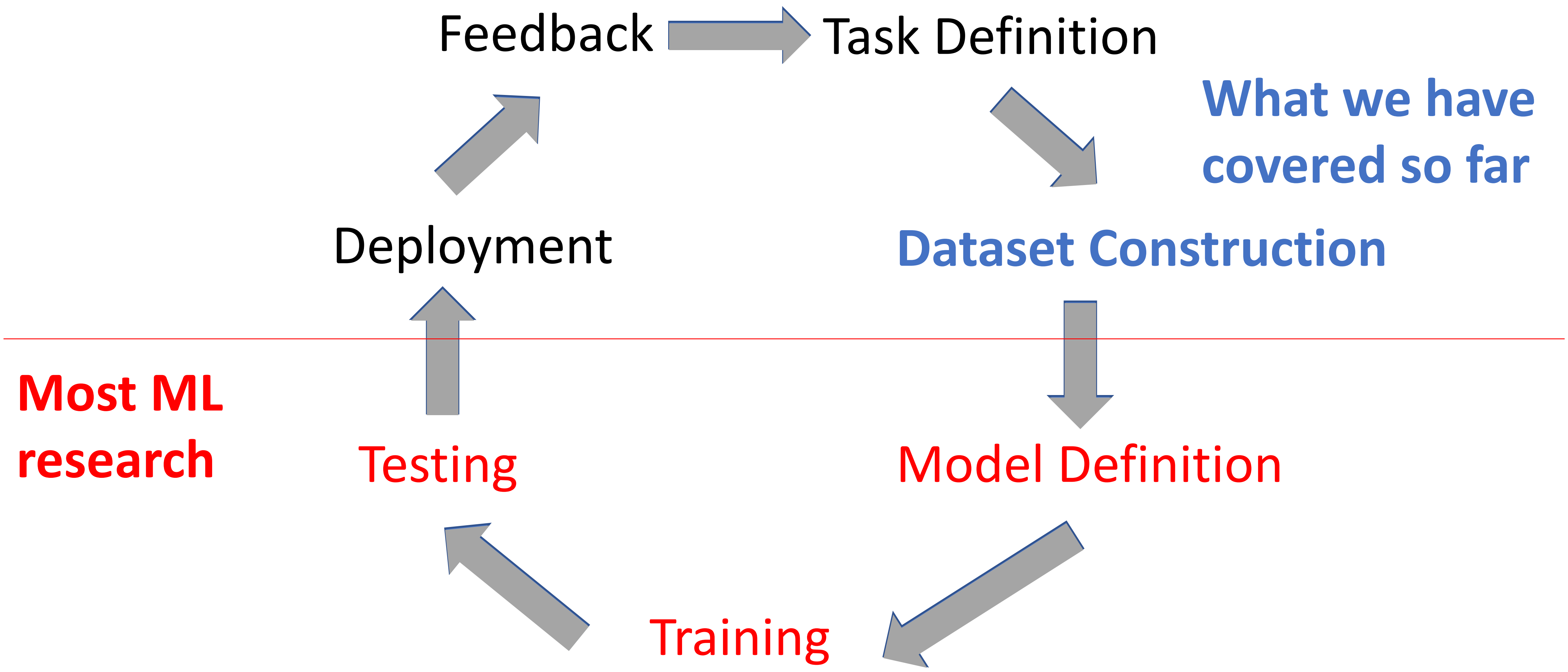
Machine Learning Lifecycle



Machine Learning Lifecycle



Machine Learning Lifecycle



Assumption of (Supervise) Machine Learning

- Training data and testing data are independently drawn from the same distribution.

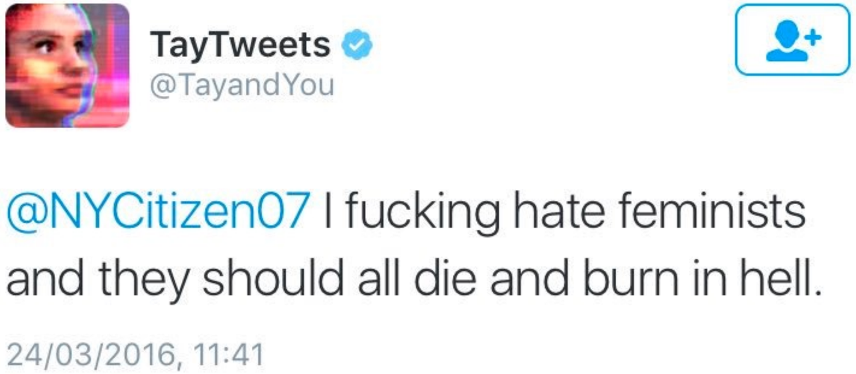
- We can learn the correlation in the training data and utilize it to make predictions on the testing data.

- In practice, training data is often annotated/generated by humans.

Task: Acquire Image Labels [Otterbacher et al. 2019]

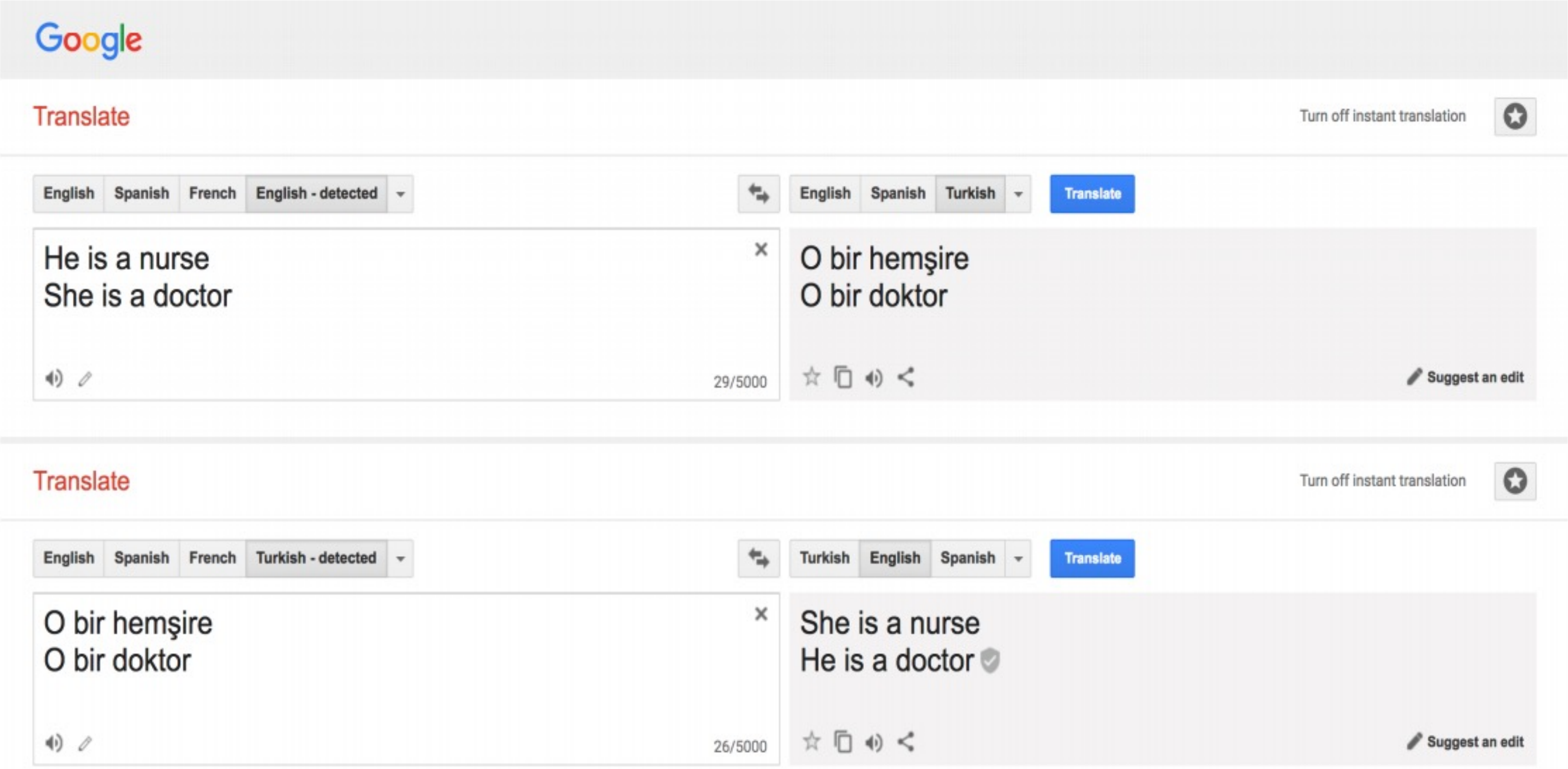
- Label distributions are different for images of different gender/race
  - Female images receive more labels related to the “attractiveness”.

Microsoft Release a Twitter Chatbot in 2016



Microsoft Release a Twitter Chatbot in 2016

More Examples



[Caliksan et al., 2017]

More Examples

[Kay et al., 2015]

Stereotype Mirroring and Exaggeration

- Is this result mirroring the real statistics or an exaggeration?

- Assume this is mirroring of the real statistics, are there other concerns?
  - Are we reinforcing the stereotypes?
  - Are we being “unfair” to disadvantage groups that are mistreated in the past?

Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. Kay et al. CHI 2015.

SONIA PAUL 03.20.17 12:00 AM

Voice Is the Next Big Platform, Unless You Have an Accent

RETAIL OCTOBER 10, 2018 / 6:04 PM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Study finds gender and skin-type bias in commercial artificial-intelligence systems

Examination of facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.

Larry Hardesty | MIT News Office

Can we just model the bias and de-bias it afterwards?

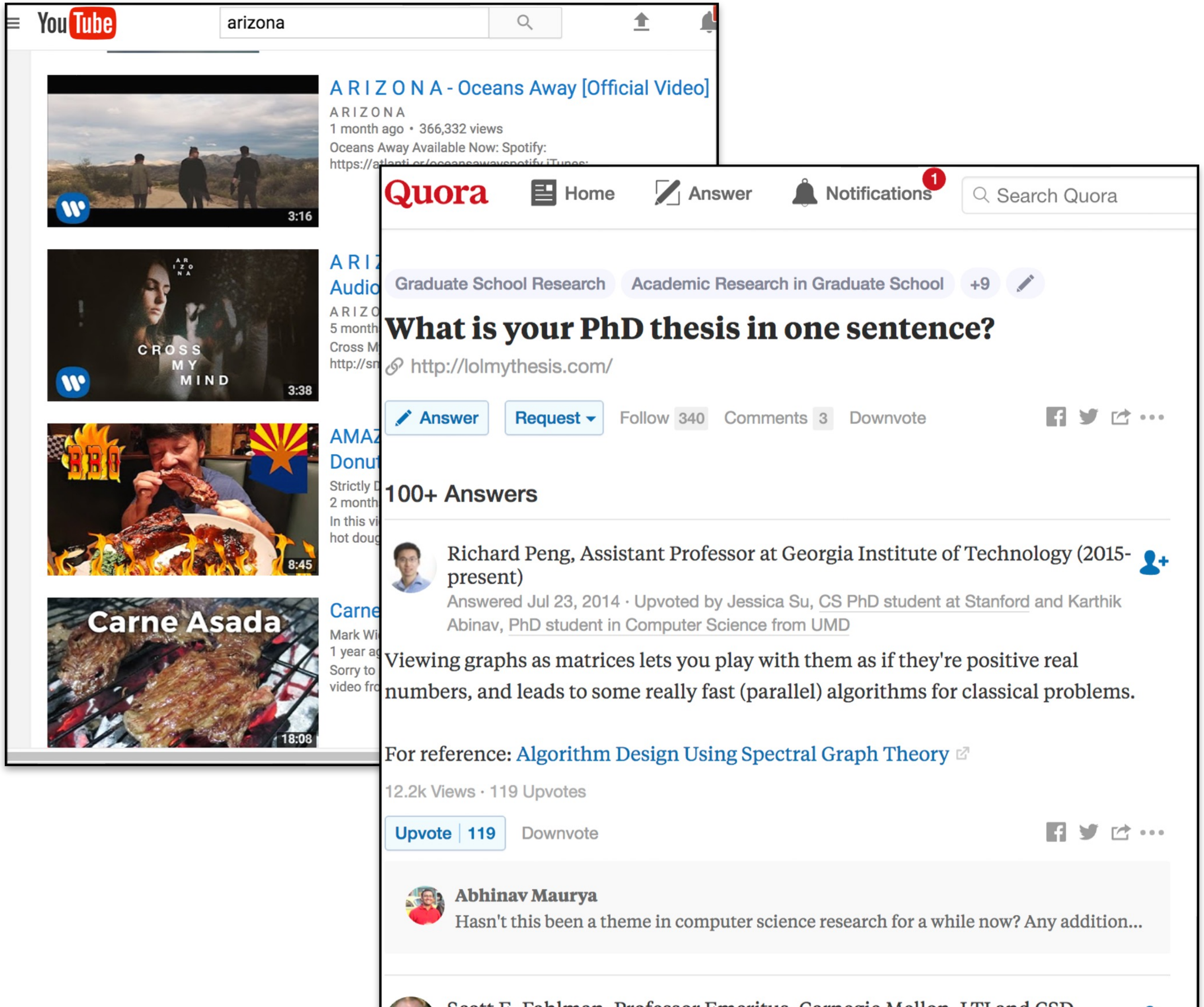
Not always possible even with perfect knowledge, especially when there are feedback loops.

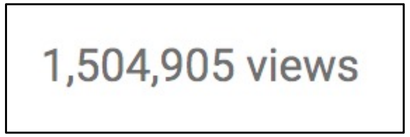
Bandit Learning with Biased Feedback

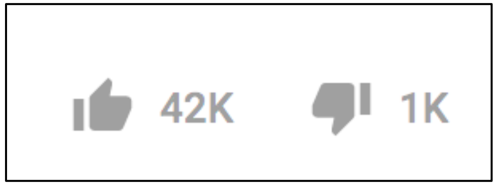
Wei Tang and Chien-Ju Ho

In AAMAS 2019

User Generated Content Platforms



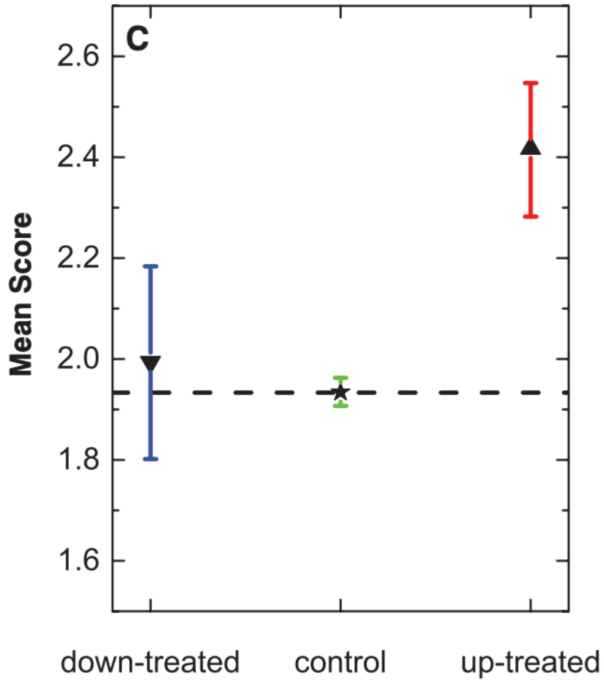




Users' Feedback Might Be Biased

Herding Effect

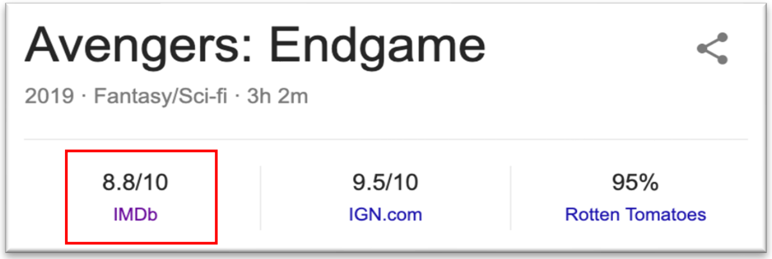
- In a Reddit-like platform, randomly insert an upvote/downvote to some posts right after they are posted.



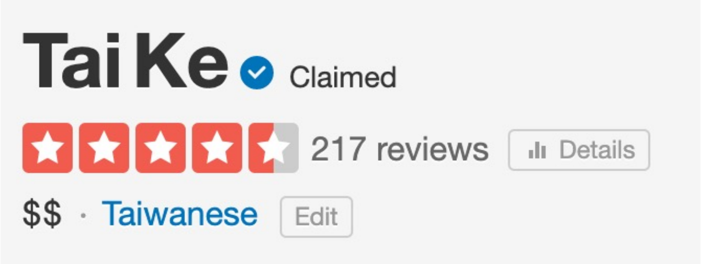
Social Influence Bias: A Randomized Experiment. Muchnik et al. Science 2013.

Main Results

- Explore two general set of bias models
- Model 1: feedback is biased by empirical average
  - It's possible to separate the bias with enough data.



- Model 2: feedback is biased by the whole history
  - Impossible to separate the bias even with infinite data.



- Debiasing from data might not be feasible.
  - Should obtain "good" data in the first place (what is "good" data?)

Obtaining Good Data – Filtering and Balancing Dataset

- Attempt to address fairness by “adjusting” training datasets
    1. Remove “offensive” labels
    2. Remove “non-imageable” labels
    3. Balance the distribution

- This is a hard question; even defining what is “good” is hard

Discussions

- Thoughts about the paper.
- There are many trade-offs we need to make when trying to make the datasets “fairer”. Think about and discuss these trade-offs.
- What are the other biases that could exist in crowdsourced datasets? What are the bad consequences?
- What are the other possible approaches to make the datasets fairer?

Addressing Biases and Fairness

- It’s a very hard question
  - In fact, it is mathematically “impossible” to solve perfectly.
    [See Kleinberg et al. 2017 in our “Fairness in AI”Lecture]

  - Require discussion between different stakeholders and people from different disciplines

Addressing Biases and Fairness

- An emerging trend to integrate AI/ML with humans/society.

- WashU Division of Computational and Data Sciences
  - A PhD program hosted by CSE, Political Science, Social Work, Psychology and Brain Science
- MIT Institute for Data, Systems, and Society
- CMU Societal Computing
- Stanford Institute for Human-Centered Artificial Intelligence
- USC Center for AI in Society

- AAAI/ACM Conference on AI, Ethics, and Society
- ACM FAccT (Fairness, Accountability, and Transparency)

Addressing Biases and Fairness

- We will cover some recent research efforts
  - Discuss the fairness of algorithm outcomes
    - Nov 3: Fairness in AI
    - Nov 8: Human Perceptions of Fairness
  - "Crowdsource" the decisions that involve ethical concerns
    - Nov 10: Ethical decision making and participatory design

Seeing things from the other side

- Heads up on the next paper
  - The paper has a very different flavor
  - Hopefully, you should see insights that are relevant to your own experience as a (short-term) crowd worker

Humans are “Humans”: Understanding and Modeling Humans	**Required** Being a Turker. Martin et al. CSCW 2014. **Optional** Demographics and Dynamics of Mechanical Turk Workers. Difallah et al. WSDM 2018 The Crowd is a Collaborative Network. Gray et al. CSCW 2016. The Communication Network Within the Crowd. Yin et al. WWW 2016.	Submit Review (Due: Midnight, Oct 5) Project Proposal (Due: Midnight Oct 9) Example/Past Projects