
Graduate Students: Project Reminder

Midpoint due is on Nov. 15 (3 weeks from now)

Midpoint presentations on Mon, Nov. 15.

Guidelines will be released this weekend

Make progress every day.

Keep a notebook & write as you go, so that you are not writing both the report and making the slides at the last minute.

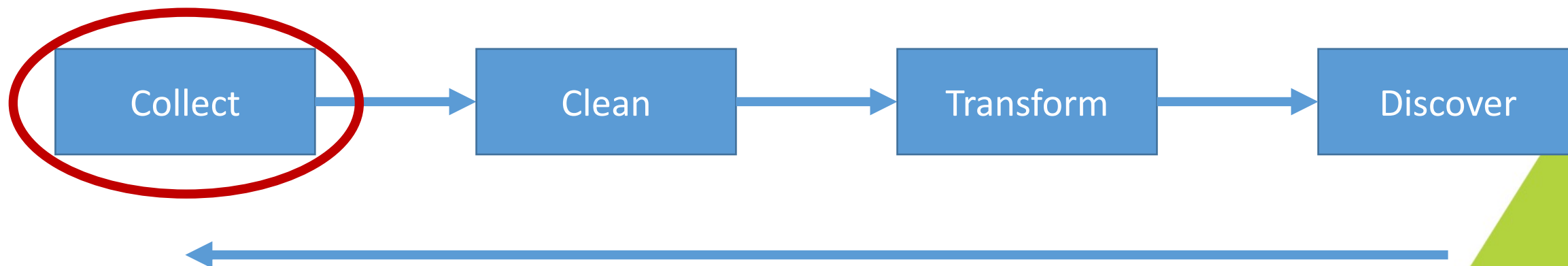
CS 295B/CS 395B
Systems for Knowledge
Discovery

Data Collection via
“Wisdom of the Crowd”



The University of Vermont

Next three weeks



Today: “wisdom of the crowds” for labeling (point estimation) tasks,
Amazon’s Mechanical Turk (AMT), RESTful APIs

Future: other tasks, ethics, role of crowdwork platforms

“Wisdom of the Crowds” for labeling


Data collection: labelling

Option A: Do the labelling yourself/in your lab (often one label per item)

Option B: Pay labelers (2-5 labels per item, use inter-annotator agreement)

Pay few experts (expensive)

Pay many non-experts (cheap)



Inter-annotator
agreement, e.g.
Cohen's kappa

Getting more out of less: Snorkle

Recall: Snorkle is a UI for efficiently acquiring *expert labels*

How (not rhetorical)?

What to do if you don't need experts?

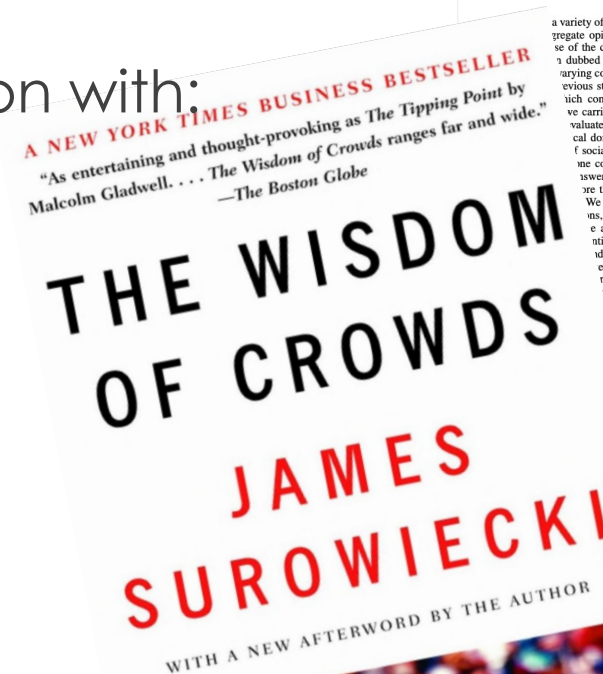
What to do if you can't find experts?

Solution: “Wisdom of the Crowd”

AKA “collective intelligence”

Concept is old, but gained new attention with:

- Web 2.0 (officially coined in 2004)
 - E.g., Wikipedia (founded 2001)
- 9/11 (2001)



Studying the “Wisdom of Crowds” at Scale

Camelia Simoiu,¹ Chiraag Sumanth,¹ Alok Mysore,² Sharad Goel¹
¹Stanford University, ²University of California San Diego

Abstract

A variety of problem domains, it has been observed that the aggregate opinions of groups are often more accurate than those of the constituent individuals, a phenomenon that has been dubbed the “wisdom of the crowd”. However, due to varying contexts, sample sizes, methodologies, and scope of previous studies, it has been difficult to gauge the extent to which conclusions generalize. To investigate this question, we carried out a large online experiment to systematically evaluate crowd performance on 1,000 questions across a wide range of domains. We further tested the effect of different social conditions on crowd performance. For example, in one condition, participants could see the cumulative answer before providing their own. In total, we collected more than 500,000 responses from nearly 2,000 participants. We have three main results. First, averaged across all questions, we find that the crowd indeed performs better than the average individual in the crowd—but we also find that initial heterogeneity in performance across questions can lead to a decrease in crowd performance. Second, we find that crowd performance is generally better than that of individuals; as a result, the crowd is able to solve more questions than any individual. Finally, we find that crowd performance is generally better than that of individuals; as a result, the crowd is able to solve more questions than any individual. Finally, we find that crowd performance is generally better than that of individuals; as a result, the crowd is able to solve more questions than any individual.

Introduction

Wise? In his 1841 book, “Memoirs of exiles and the madness of crowds,” James Surowiecki documents a series of remarkable tales of collective wisdom. From the hysteria of the South Sea bubble to the many British investors in the 1720s, from the 19th-century “tulipomania,” when investors collected tulip bulbs until a sudden crash rendered them worthless, to the value rendered them worthless decades later, in yet another classic example, Francis Galton watched as eight hundred people gathered to guess the weight of an ox at a county fair. He famously observed that the median of the guesses—1,207 pounds—was, remarkably, within 1% of the true weight (Galton 1907).

Over the past century, there have been dozens of studies that document this “wisdom of crowds” effect (Surowiecki 2005). Simple aggregation—as in the case of Galton’s ox competition—has been successfully applied to aid prediction, inference, and decision making in a diverse range of contexts. For example, crowd judgments have been used to successfully answer general knowledge questions (Surowiecki 2005), identify phishing websites and web spam (Moore and Clayton 2008; Liu et al. 2012), forecast current political and economic events (Budescu and Chen 2014; Griffiths and Tenenbaum 2006; Hill and Ready-Campbell 2011), predict sports outcomes (Herzog and Hertwig 2011; Goel et al. 2010), and predict climate-related, social, and technological events (Huefner et al. 2013; Kaplan, Skogstad, and Gershick 1950). However, given the diversity of experimental designs, subject pools, and analytic methods employed, it has been difficult to know whether these documented examples are a representative collection of a much larger space of tasks that exhibit a wisdom-of-crowds phenomenon, or conversely, whether they are highly specific instances of an interesting, though ultimately limited occurrence.

Moreover, it is unclear whether these findings generalize to many real-world settings where individuals make decisions under the influence of others’ judgments. This question is especially relevant today, as peer influence is oftentimes explicitly built into online platforms. One might choose a restaurant, watch a movie, read a news story, or purchase a book because of the aggregated opinions of the “crowd.” Recommender systems may display top-rated products first by default, whose quality has been estimated as the most popular or highly voted. In recent years, researchers have debated whether social influence undermines or enhances the wisdom of crowds. On the one hand, some have conjectured that if participants receive information about the answers of others, that can help ground responses, leading to greater accuracy (Fariu et al. 2010; King et al. 2012; Madrilas and de Polavieja 2015). But, on the other hand, there is also a worry that such social influence could result

2019

2005

Typical Example: Labeling tweets about the weather

 **cess (is studying)** @cesstwt_

There's something in the weather yesterday



0:02 34.1K views

From *7th Lilith13* 🎃❤️ *Jimtober* 🌸

9:24 PM · Oct 24, 2021 · Twitter for iPhone

 **Reggie Wayne** @ReggieWayne_17 · 10h

Halftime Humor.. Guy at the bar sees me yelling at the TV. He comes and says you must don't know how hard it is to play in **weather** like this. I looked at him and bought his next drink. Everybody looked at him like this. Hilarious!



GIF

98 382 3.3K

 **Los Angeles Times** @latimes · 13h

Bomb cyclone.

Rivers of rain.

Emergency response officials have been bracing all week for the worst days of the storm — and it finally arrived late Saturday evening.




latimes.com

Heavy rain unleashes mud, debris flows in Northern California areas b...
Heavy rain in Northern California unleashed mud and debris flows and shut down at least one critical highway by Sunday morning.

25 243 385

Typical Example: Labeling tweets about the weather

 **cess (is studying)** @cesstwt_

There's something in the weather yesterday



0:02 34.1K views

From *7th Lilith13* 🎃❤️ *Jimtober*

9:24 PM · Oct 24, 2021 · Twitter for iPhone

 **Reggie Wayne** @ReggieWayne_17 · 10h

Halftime Humor.. Guy at the bar sees me yelling at the TV. He comes and says you must don't know how hard it is to play in **weather** like this. I looked at him and bought his next drink. Everybody looked at him like this. Hilarious!



GIF

98 382 3.3K

 **Los Angeles Times** @latimes · 13h

Bomb cyclone.

Rivers of rain.

Emergency response officials have been bracing all week for the worst days of the storm — and it finally arrived late Saturday evening.



latimes.com

Heavy rain unleashes mud, debris flows in Northern California areas b...
Heavy rain in Northern California unleashed mud and debris flows and shut down at least one critical highway by Sunday morning.

25 243 385

Typical Example: Labeling tweets about the weather

 **cess (is studying)** @cesstwt_

There's something in the weather yesterday



0:02 34.1K views

From *7th Lilith13* 🎃❤️ *Jimtober*

9:24 PM · Oct 24, 2021 · Twitter for iPhone

 **Reggie Wayne** @ReggieWayne_17 · 10h

Halftime Humor.. Guy at the bar sees me yelling at the TV. He comes and says you must don't know how hard it is to play in **weather** like this. I looked at him and bought his next drink. Everybody looked at him like this. Hilarious!



GIF

98 382 3.3K

 **Los Angeles Times** @latimes · 13h

Bomb cyclone.

Rivers of rain.

Emergency response officials have been bracing all week for the worst days of the storm — and it finally arrived late Saturday evening.



latimes.com

Heavy rain unleashes mud, debris flows in Northern California areas b...
Heavy rain in Northern California unleashed mud and debris flows and shut down at least one critical highway by Sunday morning.

25 243 385

Typical Example: Labeling tweets about the weather

 **cess (is studying)** @cesstwt_

There's something in the weather yesterday



0:02 34.1K views

From *7th Lilith13* 🎃❤️ *Jimtober*

9:24 PM · Oct 24, 2021 · Twitter for iPhone

 **Reggie Wayne** @ReggieWayne_17 · 10h

Halftime Humor.. Guy at the bar sees me yelling at the TV. He comes and says you must don't know how hard it is to play in **weather** like this. I looked at him and bought his next drink. Everybody looked at him like this. Hilarious!



GIF

98 382 3.3K

 **Los Angeles Times** @latimes · 13h

Bomb cyclone.

Rivers of rain.

Emergency response officials have been bracing all week for the worst days of the storm — and it finally arrived late Saturday evening.



latimes.com

Heavy rain unleashes mud, debris flows in Northern California areas b...
Heavy rain in Northern California unleashed mud and debris flows and shut down at least one critical highway by Sunday morning.

25 243 385

2/3 Game?

Guess the number that is $\frac{2}{3}$ of the average of everyone's guesses.

Guesses must be between 0 and 100, inclusive.

```
mirror_mod = modifier_ob.  
set mirror object to mirror.  
mirror_mod.mirror_object  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True  
  
#selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob.  
mirror_ob.select = 0  
= bpy.context.selected_object  
data.objects[one.name].select  
  
print("please select exactly  
  
-- OPERATOR CLASSES ----  
  
types.Operator):  
on X mirror to the selected  
object.mirror_mirror_x"  
mirror X"  
  
context):  
context.active_object is not
```

This is a recurring theme

What are some examples?

- Labeling
- Machines
- Databases
- Machine learning

Consider implications for average case vs. worst case

Point estimation tasks

Automan paper focuses on point estimation

- Looking for convergence to a single estimate
- Quantities estimated are typically categorical variables

Why does this matter?

Data collected is “work” → contentious interpretation

Definitely not true of other data

Crowdsourcing platforms

Crowdsourcing

Process of collecting data from a large group.

Typically non-experts.

Typically facilitated by *crowdsourcing platforms*.

Crowdsourcing Platforms for Collection

Platform	Domain	Requester Access Model

Crowdsourcing Platforms for Collection

Platform	Domain	Requester Access Model
Amazon's Mechanical Turk (AMT)	Any virtual task	Ad hoc with Amazon Account, commission-based

Crowdsourcing Platforms for Collection

Platform	Domain	Requester Access Model
Amazon's Mechanical Turk (AMT)	Any virtual task	Ad hoc with Amazon Account, commission-based
CrowdFlower/Figure Eight	Any virtual task	Subscription access

Crowdsourcing Platforms for Collection

Platform	Domain	Requester Access Model
Amazon's Mechanical Turk (AMT)	Any virtual task	Ad hoc with Amazon Account, commission-based
CrowdFlower/Figure Eight	Any virtual task	Subscription access
ClickWorker	AI training data	Tiered (commission → subscription)

Crowdsourcing Platforms for Collection

Platform	Domain	Requester Access Model
Amazon's Mechanical Turk (AMT)	Any virtual task	Ad hoc with Amazon Account, commission-based
CrowdFlower/Figure Eight	Any virtual task	Subscription access
ClickWorker	AI training data	Tiered (commission → subscription)
RapidWorkers	Any virtual task	Unclear

Amazon's Mechanical Turk (AMT)

Most common/popular crowdsourcing system. Why?

- Does not need to be profitable
- Anti-competitive behavior & account integration
 - Low barrier to entry for requesters
- Ease of use
 - Integrated with developer tools
 - multiple APIs, documentation, sandboxing, templates

Myth of Excess Capacity

Idea: Do some quick work for pennies while, e.g. waiting in line

Next week: *Who is actually doing these jobs?*

Broader “excess capacity” trends: AirBnB, rideshares, gig economy

December: *How to design sociotechnical systems for knowledge discovery when these systems inhabit and mutate a broader social context?*

RESTful APIs

Refresher: APIs

API == Application Programming Interface

How two programs communicate

Variable scale of communication:

- Within programs
- Across the internet

Within programs – can send encoded objects

Across the internet – need to think about encoding data

Refresher: Data Exchange

Sending data across applications or the internet

- Always just sending bytes
- Assume we have a chunk of bytes we decode as a string of data
 - Represent this data in a structured way
 - JSON, XML, ProtoBuf, etc.

Can always have two programs locally communicating via files, sockets, etc.

Data exchange formats make parsing into objects easy!

HTTP

Over the internet: wrap in a HTTP *request*

Important requests:

GET – *supposed to be read-only (like DML)*

PUT – *supposed to be idempotent lightweight modification to server (like DDL)*

POST – *heavyweight data, possible modification to server (like DDL)*



Encoded in URLs

SEARCH DIRECTORY

Name or NetID:

Search for the contact information for all UVM students, faculty and staff, by entering a name (partial or full) or a UVM netID.

[Common UVM Contacts](#)
[Update My Directory Information](#)

Advanced Options ▾

SEARCH RESULTS

Name	Phone Number	Email	NetID	Department / Year
Emma Fox		Emma.Fox@uvm.edu	efox	Junior
Emma Rose Abel		Emma.Abel@uvm.edu	eabel	Junior
Emma Lavella Eash		Emma.Eash@uvm.edu	eeash	First Year
Emma K Gwyn	+1 802 656 2975	Emma.Gwyn@uvm.edu	egwyn	Rubenstein Sch Env & Nat Res
Emma L Huse		Emma.Huse@uvm.edu	ehuse	International Educ Svcs Admin
Emma Zhao	+1 802 847 2259	Emma.Zhao@uvm.edu	ezhao	Psychiatry
Grace Emma Clay		Grace.Clay@uvm.edu	gclay	Senior
Jemma N Hoko		Jemma.Hoko@uvm.edu	jhoko	Sophomore
Emma A Page		Emma.Page@uvm.edu	eapage	Continuing Education Student
Emma Anne Rose Askew		Emma.Askew@uvm.edu	easkew	Senior
Emma C Baird	+1 802 656 3376	Emma.Baird@uvm.edu	ebaird	Social Work Outreach

HTTP Example

Example GET request

Theoretically can do this in Python:

```
import requests
reply = requests.get(base_url,
                    params={'search': 'emma'})
```

(Actually need to set some headers)

<https://www.uvm.edu/directory?search=emma&n=5>

REST API

REST == *Representational state transfer*

Not a specification or protocol, but architectural best practices

Important features from the client-side programmer's perspective:

- Separation of concerns
- Statelessness

Why do we like RESTful APIs?

API → interaction can be programmatic

RESTful → interaction is predictable

Fixed data exchange format → can work entirely within your IDE!

But...people are not APIs!