

Crowdsourcing

Introduction, Crowd Computing, Creative Applications

Malay Bhattacharyya

Assistant Professor
Machine Intelligence Unit
Indian Statistical Institute, Kolkata

Guest Lecture at IIT, Kharagpur (October 03, 2018)





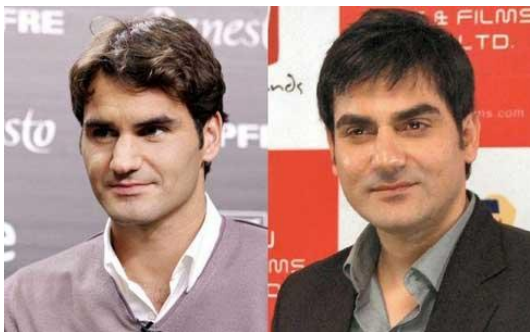


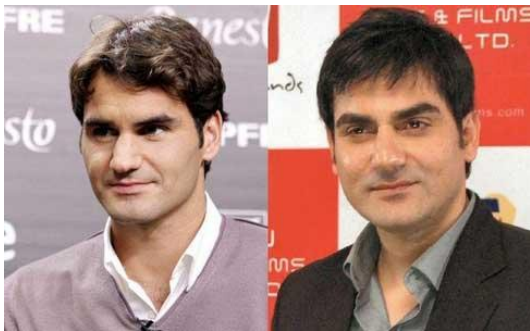
Ranbir Kapoor



Ranbir Kapoor

Same person may appear to be different!!!





Roger Federer looks like Arbaaz Khan!!!

Different persons may appear to be the same!!!

So, human power ('perception/learning' in the previous two examples) is still better than the computational power for some complex tasks like face recognition.

So, human power ('perception/learning' in the previous two examples) is still better than the computational power for some complex tasks like face recognition.

What if we can combine the powers of human and computation together? But how?

So, human power ('perception/learning' in the previous two examples) is still better than the computational power for some complex tasks like face recognition.

What if we can combine the powers of human and computation together? But how?

Human Computation

Introduction

Crowd Computing

Creative Applications

References

Background

Crowdsourcing: The power of many replacing a specialized few.

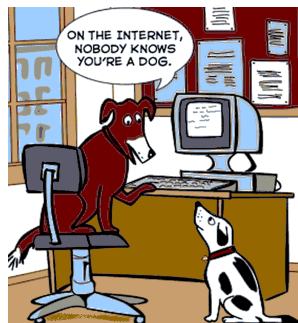
Background

Crowdsourcing: The power of many replacing a specialized few.



Background

Crowdsourcing: The power of many replacing a specialized few.



Definition (Crowdsourcing [Brabham, 2013])

Crowdsourcing is an online, distributed problem-solving and production model.

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

2006: The term ‘crowdsourcing’ is coined by Jeff Howe

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

2006: The term ‘crowdsourcing’ is coined by Jeff Howe

2006: Games with a purpose (ESP) [Ahn, 2006]

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

- 1714:** The Longitude Prize offered by the British government
- 1884:** Cataloging words by Oxford English Dictionary
- 2005:** Online crowd-powered system Amazon Mechanical Turk
- 2006:** The term ‘crowdsourcing’ is coined by Jeff Howe
- 2006:** Games with a purpose (ESP) [Ahn, 2006]
- 2006:** Crowd-powered biotechnology company 23andMe

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

- 1714:** The Longitude Prize offered by the British government
- 1884:** Cataloging words by Oxford English Dictionary
- 2005:** Online crowd-powered system Amazon Mechanical Turk
- 2006:** The term ‘crowdsourcing’ is coined by Jeff Howe
- 2006:** Games with a purpose (ESP) [Ahn, 2006]
- 2006:** Crowd-powered biotechnology company 23andMe
- 2007:** Citizen science company Zooniverse

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

- 1714:** The Longitude Prize offered by the British government
- 1884:** Cataloging words by Oxford English Dictionary
- 2005:** Online crowd-powered system Amazon Mechanical Turk
- 2006:** The term ‘crowdsourcing’ is coined by Jeff Howe
- 2006:** Games with a purpose (ESP) [Ahn, 2006]
- 2006:** Crowd-powered biotechnology company 23andMe
- 2007:** Citizen science company Zooniverse
- 2008:** The crowd-powered learning started with WikiProjects

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

2006: The term ‘crowdsourcing’ is coined by Jeff Howe

2006: Games with a purpose (ESP) [Ahn, 2006]

2006: Crowd-powered biotechnology company 23andMe

2007: Citizen science company Zooniverse

2008: The crowd-powered learning started with WikiProjects

2009: DARPA’s Red Balloon contest [Tang, 2011], Netflix Prize

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

2006: The term ‘crowdsourcing’ is coined by Jeff Howe

2006: Games with a purpose (ESP) [Ahn, 2006]

2006: Crowd-powered biotechnology company 23andMe

2007: Citizen science company Zooniverse

2008: The crowd-powered learning started with WikiProjects

2009: DARPA’s Red Balloon contest [Tang, 2011], Netflix Prize

2013: Crowdsourcing efforts for street-level accessibility

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

1714: The Longitude Prize offered by the British government

1884: Cataloging words by Oxford English Dictionary

2005: Online crowd-powered system Amazon Mechanical Turk

2006: The term ‘crowdsourcing’ is coined by Jeff Howe

2006: Games with a purpose (ESP) [Ahn, 2006]

2006: Crowd-powered biotechnology company 23andMe

2007: Citizen science company Zooniverse

2008: The crowd-powered learning started with WikiProjects

2009: DARPA’s Red Balloon contest [Tang, 2011], Netflix Prize

2013: Crowdsourcing efforts for street-level accessibility

2016: Crowd-powered surveillance

History of crowdsourcing

“With the crowds on your side, it’s easier to play up to your potential”

– Julius Erving.

- 1714:** The Longitude Prize offered by the British government
- 1884:** Cataloging words by Oxford English Dictionary
- 2005:** Online crowd-powered system Amazon Mechanical Turk
- 2006:** The term ‘crowdsourcing’ is coined by Jeff Howe
- 2006:** Games with a purpose (ESP) [Ahn, 2006]
- 2006:** Crowd-powered biotechnology company 23andMe
- 2007:** Citizen science company Zooniverse
- 2008:** The crowd-powered learning started with WikiProjects
- 2009:** DARPA’s Red Balloon contest [Tang, 2011], Netflix Prize
- 2013:** Crowdsourcing efforts for street-level accessibility
- 2016:** Crowd-powered surveillance
- 2018:** Finding American soldiers who served in WWII

Categories of crowdsourcing

- **Based on designs:** Competitive and collaborative.

Categories of crowdsourcing

- **Based on designs:** Competitive and collaborative.
- **Based on demands:** Real-time and non real-time.

Categories of crowdsourcing

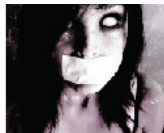
- **Based on designs:** Competitive and collaborative.
- **Based on demands:** Real-time and non real-time.
- **Based on motivations:** Solution finding, opinion seeking, content creation, design competitions, data collection, manual tasks, testing, customer service, programming, crowdfunding.

Categories of crowdsourcing

- **Based on designs:** Competitive and collaborative.
- **Based on demands:** Real-time and non real-time.
- **Based on motivations:** Solution finding, opinion seeking, content creation, design competitions, data collection, manual tasks, testing, customer service, programming, crowdfunding.
- **Based on applications:** Crowdvoting, wisdom of the crowd, crowdfunding, microwork, creative crowdsourcing and inducement prize contests.

Getting the crowd – Amazon Mechanical Turk (MTurk)

Which is scarier?



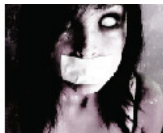
Which person looks more heroic?



Nature of jobs posted on MTurk

Getting the crowd – Amazon Mechanical Turk (MTurk)

Which is scarier?



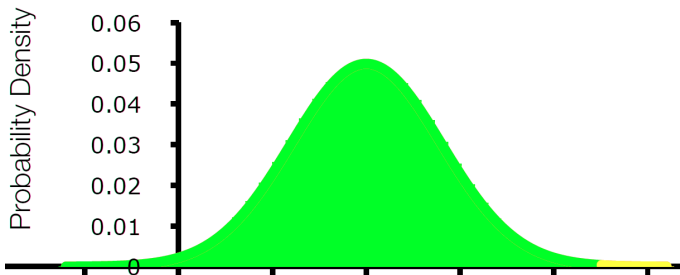
Which person looks more heroic?



Nature of jobs posted on MTurk

Note: The jobs are posted in the form of Human Intelligence Task (HITs) on MTurk.

Working principle – ‘Extreme value outcomes’



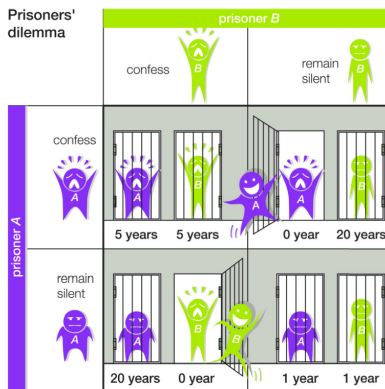
Crowdsourcing enables the exploration of ‘extreme value outcomes’ from numerous submissions (not necessarily involving the experts)

The ESP game

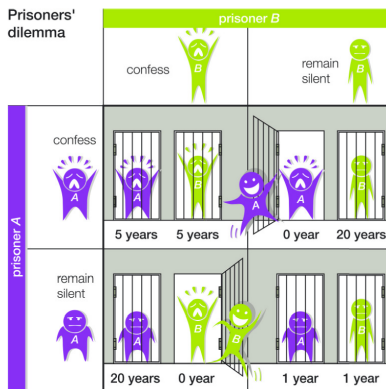


Through online games, people can collectively solve large-scale computational problems [Ahn, 2006].

Prisoners' dilemma



Prisoners' dilemma



$N = \{1, 2\}$; $S_1 = S_2 = \{C, S\}$; $u_1(C, C) = -5$, $u_2(C, C) = -5$,
 $u_1(C, S) = 0$, $u_2(C, S) = -20$, $u_1(S, C) = -20$, $u_2(S, C) = 0$,
 $u_1(S, S) = -1$, $u_2(S, S) = -1$, where C and S denote 'confess'
 and 'remain silent', respectively.

The reCAPTCHA

Select all images with **coffee**.

Report a problem

Verify

DARPA Red Balloon Challenge



The challenge of finding coordinates of some balloons

DARPA Red Balloon Challenge



The solution of DARPA Red Balloon Challenge

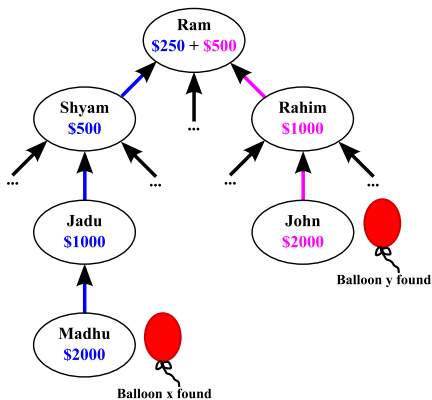
DARPA Red Balloon Challenge



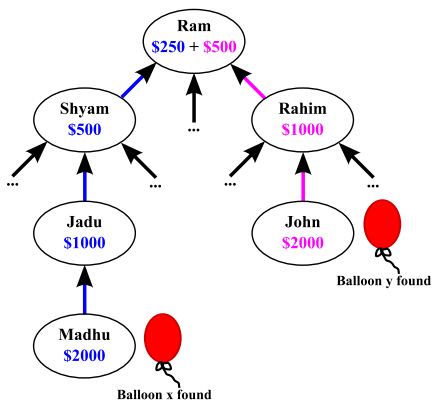
The solution of DARPA Red Balloon Challenge

Note: The best solution of DARPA Red Balloon Challenge was found in less than 9 hours.

DARPA Red Balloon Challenge – Winning strategy



DARPA Red Balloon Challenge – Winning strategy



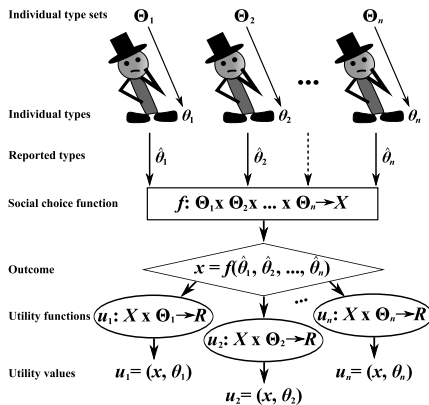
So, the maximum cost of finding a balloon becomes

$$\$2000 + \$1000 + \$500 + \dots$$

$$= \$2000\left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = \$4000 \text{ [Tang, 2011].}$$

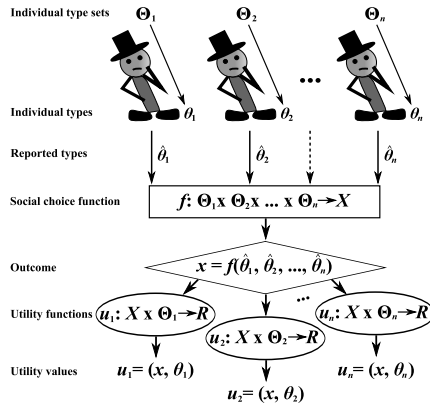
Mechanism design environment (MDE)

A standard view of MDE is shown below.



Mechanism design environment (MDE)

A standard view of MDE is shown below.



The main tasks are – preference elicitation and preference aggregation.

Word Processing

CONCLUSION

The following conclusion was Shortn'ed to 85% length:

This paper presents Soylent, a word processing interface that uses crowd workers to help with proofreading, document shortening, editing and commenting tasks. Soylent is an example of a new kind of interactive user interface in which the end user has direct access to a crowd of workers for assistance with tasks that require human attention and common sense. Implementing these kinds of interfaces requires new software programming patterns for interface software, since crowds behave differently than computer systems. We have introduced one important pattern, Find-Fix-Verify, which splits complex editing tasks into a series

Word Processing

CONCLUSION

The following conclusion was Shortn'ed to 85% length:

This paper presents SoyLent, a word processing interface that uses crowd workers to help with proofreading, document shortening, editing and commenting tasks. SoyLent is an example of a new kind of interactive user interface in which the end user has direct access to a crowd of workers for assistance with tasks that require human attention and common sense. Implementing these kinds of interfaces requires new software programming patterns for interface software, since crowds behave differently than computer systems. We have introduced one important pattern, Find-Fix-Verify, which splits complex editing tasks into a series

Soylent: A crowd-powered word processor that integrates paid crowd workers from MTurk [Bernstein, 2010]

Word Processing

Features of Soylent:

- Shortn: Document summarization of selected text down to 85% of its original length typically without changing the meaning of the text or introducing errors.
- Crowdproof: Spelling and grammar checking to find problems other word processing tools miss, explains the problems, and suggests fixes.
- The Human Macro: Offloading arbitrary word processing tasks such as formatting citations or finding appropriate figures.

Soylent improves worker quality through the Find-Fix-Verify crowd programming pattern, which splits tasks into a series of generation and review stages.

Scalable annotation

Data is the new oil

Scalable annotation

Data is the new oil, but it is still crude ...

Scalable annotation

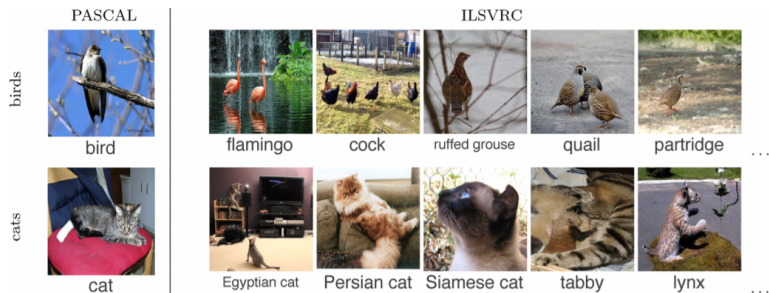
Data is the new oil, but it is still crude ...

For testing scalable Machine Learning algorithms, we need large-scale annotated (Gold) data.

Scalable annotation

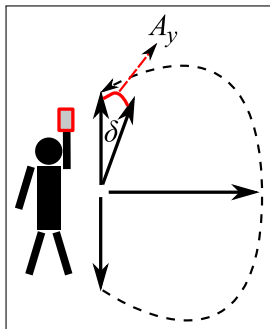
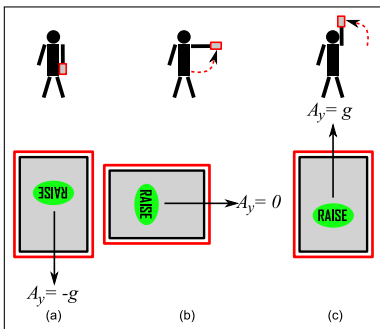
Data is the new oil, but it is still crude ...

For testing scalable Machine Learning algorithms, we need large-scale annotated (Gold) data.



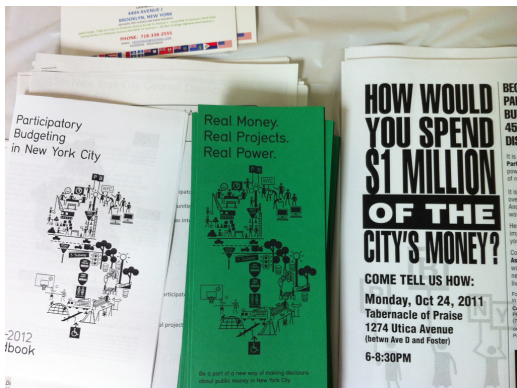
ImageNet Large Scale Visual Recognition Challenge
[Russakovsky, 2015]

Mobile healthcare

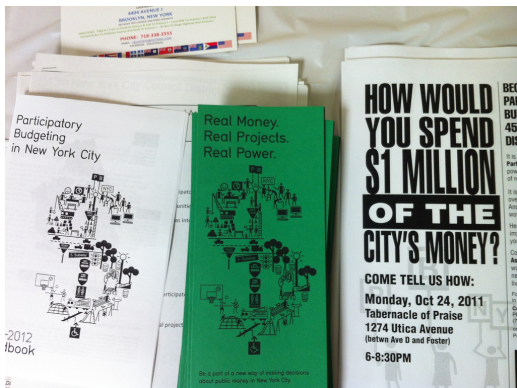


Facial, arm, and speech weakness test (FAST) for stroke recognition

Participatory budgeting



Participatory budgeting



Participatory budgeting is an approach of democratic decision-making to allocate a part of Government budget based on the opinions received from common people (crowd).

References

-  L. V. Ahn (2006) *Computer*, 39(6):92-94.
-  D. C. Brabham (2013) *Crowdsourcing*. MIT Press, Cambridge, MA, USA.
-  M. S. Bernstein *et al.* (2010) *UIST*, ACM Press, 313-322.
-  O. Russakovsky (2015) *International Journal of Computer Vision*, 115(3):211-252.
-  J. C. Tang *et al.* (2011) *CACM*, 54(4):78.

That's all for today!!!