MAKING SMART SOCIETIES FAIR

Michael Rovatsos, University of Edinburgh

Computational Decision Making and Data Science Workshop
12th July 2017, Belgrade

COLLECTIVE INTELLIGENCE

- Digital interaction platforms enable massive-scale collaboration
  - Social media
  - Crowdsourcing
  - Human computation
  - Sharing economy
- "Smartness" stems from combination of human and machine capabilities
- Can only be achieved if humans engage and are treated fairly

SMARTSOCIETY

- Hybrid and Diversity-Aware Collective Adaptive Systems: "*when people meet machines to build a smarter society*"
- €6.8M FET Integrated Project, co-ordinated by University of Trento (2013-2016)
- Brought together AI, computer science, human factors, privacy, ethics

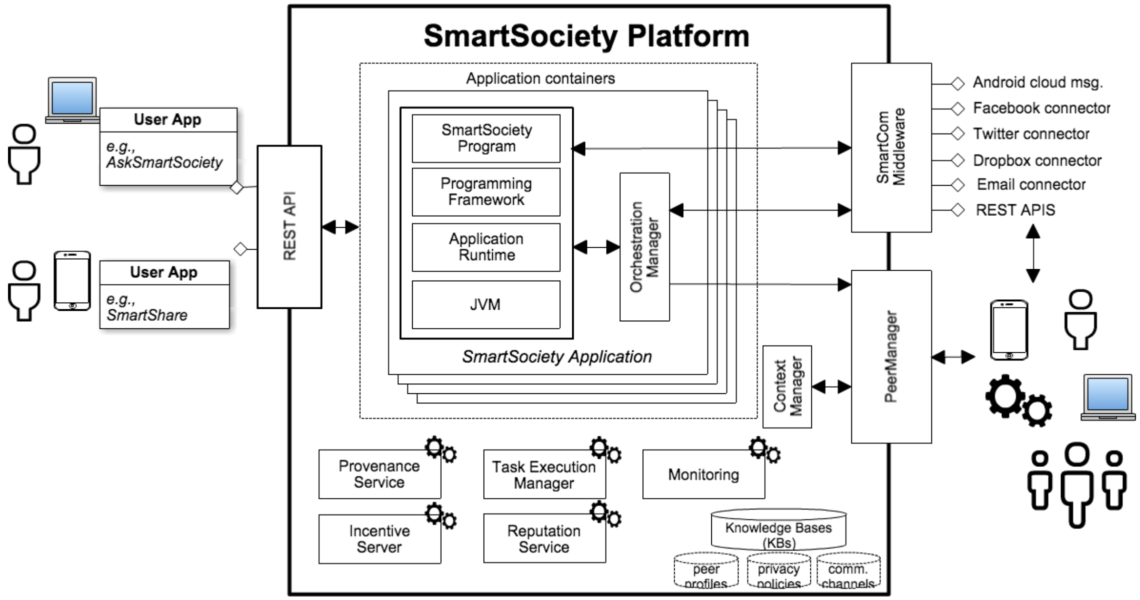
RIDESHARING

- An example sharing economy domain for smart platforms
- Involves human focus with combinatorial computational problem
- Focus on providing technology that addresses diversity among users

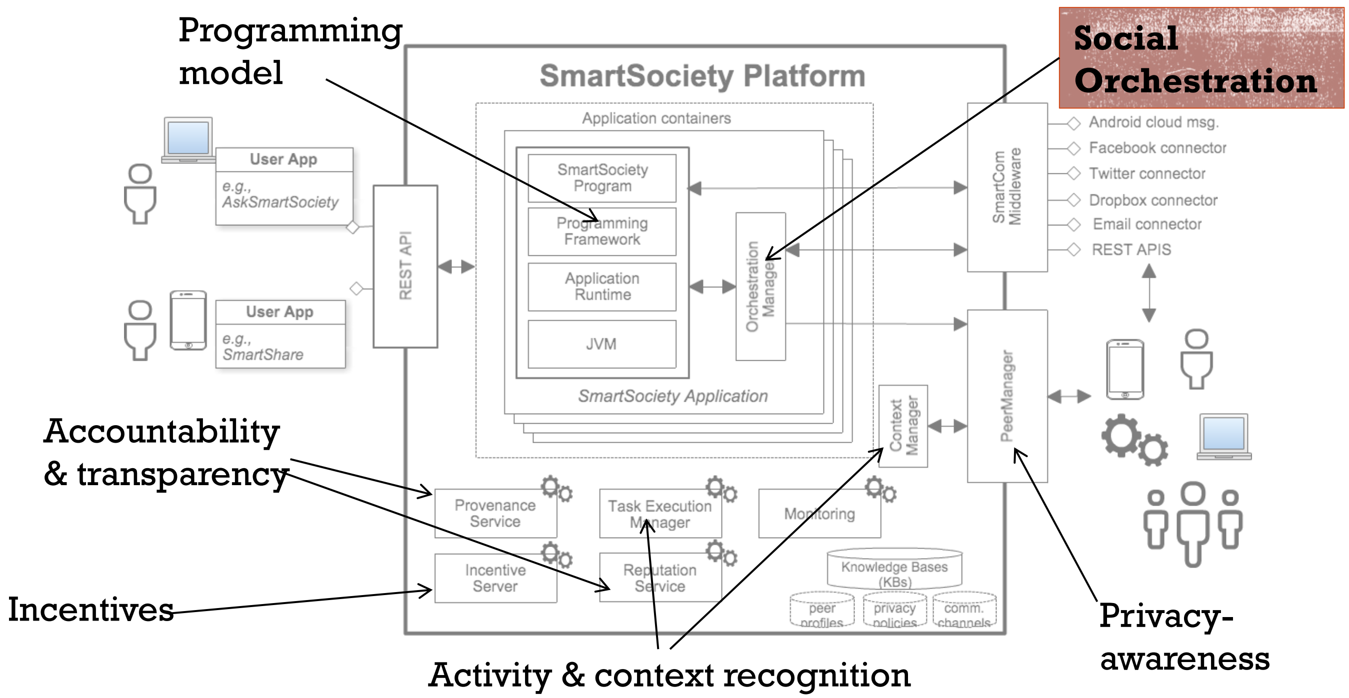
ASK & SHARE

- Doing things together **without** knowing the what/who/how
- Combines human-driven crowdsourcing with machine-driven activity recommendation
- A first step toward general collective human-machine problem solving

THE SMARTSOCIETY PLATFORM



THE SMARTSOCIETY PLATFORM



SOCIAL ORCHESTRATION

SOCIAL ORCHESTRATION

TASK COMPOSITION

- Combinatorial problem of allocating **groups** of users to **shared** tasks, where **task requests** come from users
- **Hard constraints** restrict the groupings and task properties that can be realised in principle
- **Soft constraints** determine which coalition structures and task features are preferred by system and/or users

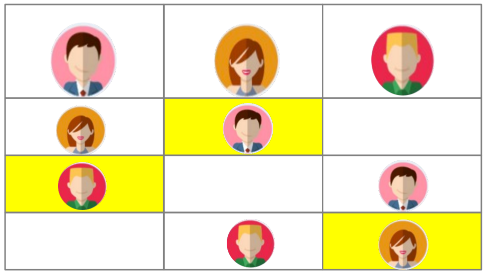
DIVERSITY VS. TASK COMPOSITION

- In traditional mechanism design, global allocations are computed given individual preferences and global criteria
  - E.g. social welfare maximisation, Pareto optimality, strategy-proofness, etc.
- Mechanisms are proposed that provably satisfy these properties, solution can therefore be **imposed** on users
- **Diversity** implies that users cannot report their preferences
  - System never captures all relevant decision variables
  - Solutions cannot be computed/considered exhaustively
  - Utility of solutions cannot be determined by users a priori

TASK ALLOCATION VS TASK RECOMMENDATION

Key problems:

- 1. How to compute "optimal" **sets** of solutions
- 2. How to **influence** users' choices
- 3. How to **learn** users' preferences



OPTIMALITY CRITERIA

- User's utility function u_i depends on user's requirements and preferences
- Global (system) utility function depends on social welfare and maximal task completion

$$U_s = \sum_{i \in I} u_i + \sum_{i \in I} \sum_{j \in J} x_{i,j}$$

COMPUTING ALLOCATIONS

MIP* $\rightarrow V^*$	**Objective**	$max_{a \in A} U_s(a)$
	Constraints	Hard feasibility constraints
MIP$^{\text{first}}$ $\rightarrow a \in R$	**Objective**	$min_{a \in A} \sum_{i \in I} \sum_{i' \in I \mid i' > i} \lvert u_i(a) - u_{i'}(a) \rvert$
	Constraints	MIP* $U_s(a) \cdot h \geq V^*$
MIP$^{\text{others}}$ $\rightarrow a' \in R$	**Objective**	$min_{a' \in A} \sum_{i \in I} \lvert u_i(a) - u_i(a') \rvert$
	Constraints	MIP$^{\text{first}}$ $a' \notin R$

INFLUENCING USERS

- We want to modify users' utility artificially so that their choices lead to a feasible global solution
- Explicit Approaches:
  - intervention
  - (possible) future reward
- Implicit Approaches:
  - discounts
  - **taxation**

TAXATION SCHEME

MIP*

$\rightarrow V^*$

Sponsored Solution

MIP$^{\text{first}}$

$\rightarrow a \in R$

MIP$^{\text{others}}$

$\rightarrow a' \in R$

Objective

$$min \sum_{i \in I} |u_i(a) - u_i(a') + \tau_i(a')| + M\Big(\sum_{i \in I} \big(u_i(a) + \epsilon - u_i(a) + \tau_i(a')\big)\Big)$$

Constraints MIP$^{\text{first}}$ $a' \notin R$

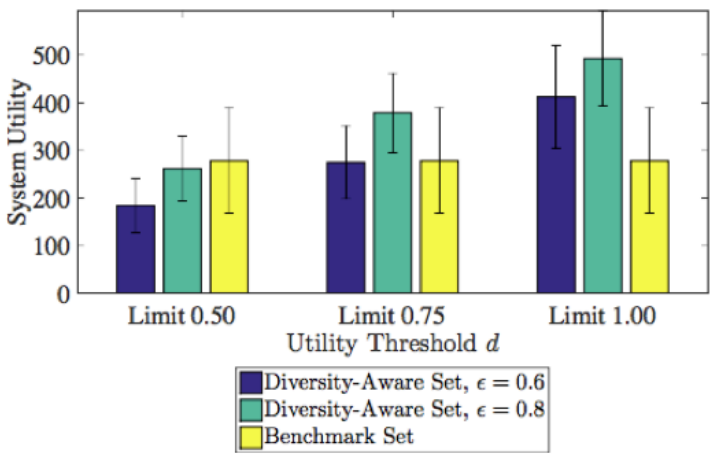
Noiseless and Constant Noise Models

$$u_i(a) + \epsilon \geq u_i(a') - \tau_i(a')$$

Logit Model (also goes into objective function)

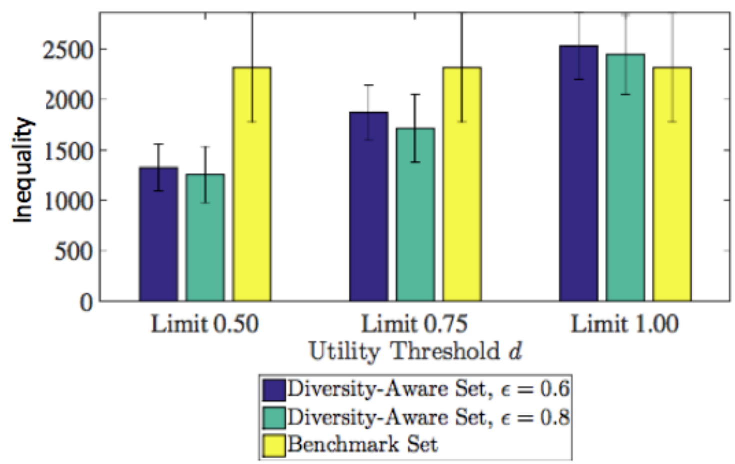
$$\frac{u_i(a)}{\Big(\sum_{a'' \in R} \big(u_i(a'') - \tau_i(a'')\big) + u_i(a') - \tau_i(a')\Big)} \geq \psi$$

RESULTS



System Utility
0
100
200
300
400
500
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set

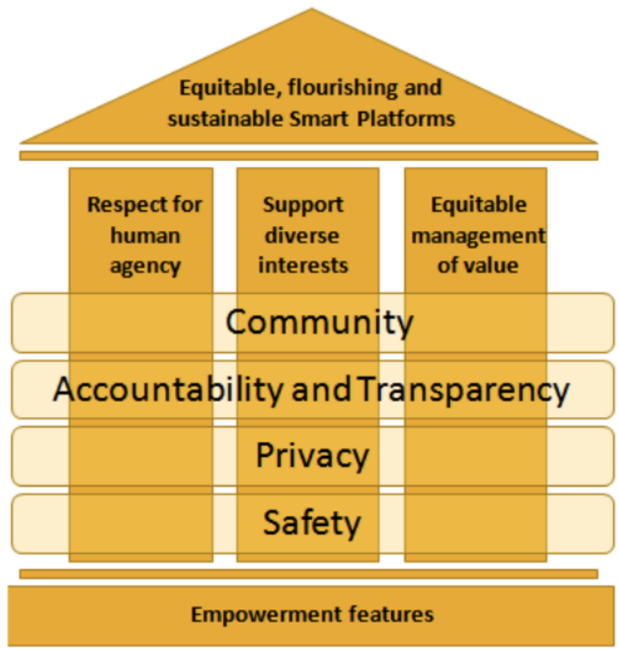
(a)



Inequality
0
500
1000
1500
2000
2500
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set

(b)

THE SOCIAL CHARTER



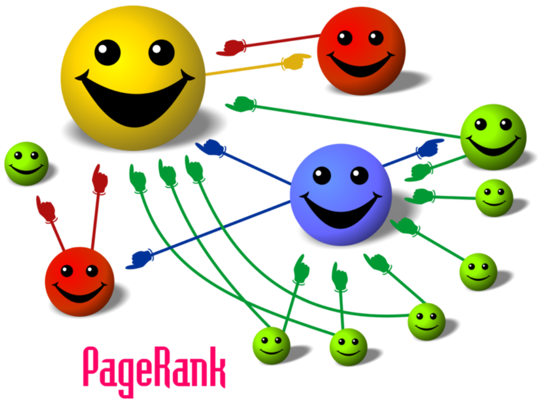
UNBIAS

- UnBias: Emancipating Users Against Algorithmic Biases for a Trusted Digital Economy
- £1.1M 2-year project led by Nottingham with Edinburgh and Oxford
- Focus on young people's perception of how their lives are influenced by algorithms

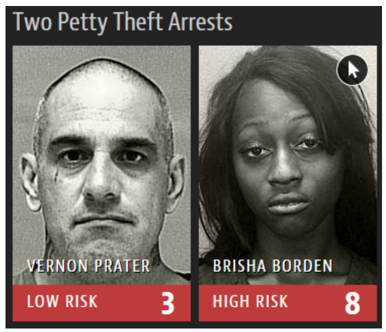
RISKS

TayTweets
@TayandYou
@BASED_AN0N Jews did 9/11.
Gas the kikes- race war now!!!
#KKK
assisted
SERENDIPITY
Get notified when the scales of love tip in your favor at your favorite local hangouts!
Your Favorite Bar
Google
white people stole my car
Maps News Shopping
Web
Show options...
Did you mean: black people stole my car
WIKIPEDIA
The Free Encyclopedia
Mary Jo Kopechne
From Wikipedia, the free encyclopedia
Editing of this article by new or unregistered users is currently disabled until August 29, 2009 due to vandalism.
Mary Jo Kopechne (July 26, 1940 – July 18, 1969) was an American teacher, secretary and political campaign specialist who died in a car accident on Chappaquiddick Island while being driven by United States Senator Ted Kennedy.
Contents
1 Life
2 Death
3 References
Mary Jo Kopechne

THE ROLE OF ALGORITHMS

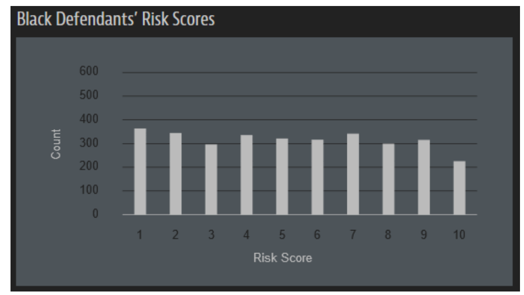


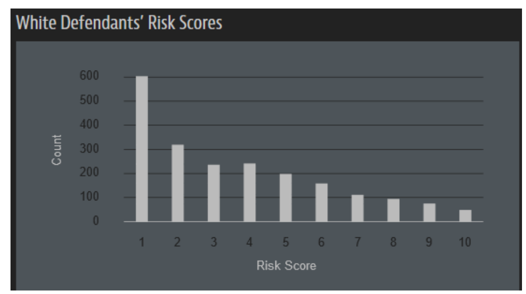
THE ROLE OF DATA



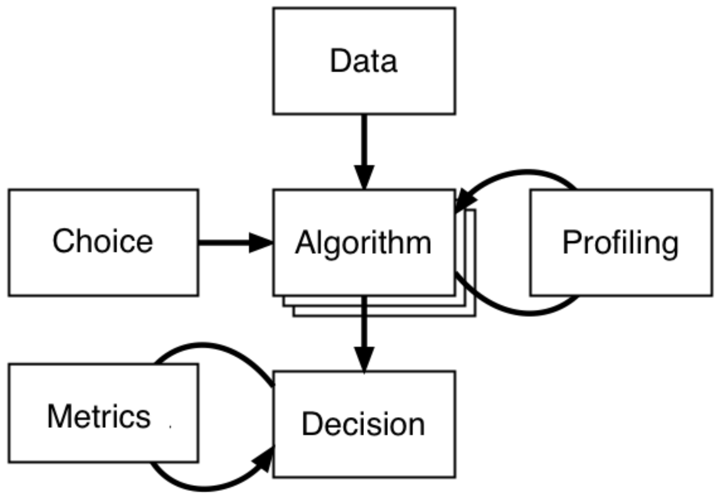
Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%





MAKING ALGORITHMS FAIR



CRITERIA: GAME THEORY

- **Social Welfare**: The solution maximises the sum of the utilities of all agents.
- **Equity**: The solution minimises the difference between all agents' individual utilities among all possible solutions.
- **MaxiMin**: The solution maximises the utility of the agent who is worst off.
- **Monotonicity**: If a different outcome is produced when agents change their preferences, it must be because at least one player benefits from that.

CRITERIA: MACHINE LEARNING

- **Unawareness:** An algorithm is fair so long as any sensitive attributes are not explicitly used in the decision-making process.
- **Rawlsian fairness:** Those who are at the same level of ability, and have the same willingness to use them, should have the same prospect of success regardless.
- **Individual fairness:** Algorithm provides similar outcomes for similar individuals, ignoring their protected attributes.
- **Demographic parity:** An algorithm is fair if its predictions are independent of the sensitive attributes across sub-populations.

CRITERIA: MACHINE LEARNING

- **Equal Opportunity:** An algorithm is fair if it is equally accurate for each value of the sensitive attributes.
- **Equalized Odds:** An algorithm is fair if it is equally accurate for each value of the sensitive attributes, for each values of the non-sensitive attributes.
- **Counterfactual fairness:** A decision is fair toward an individual if it gives the same predictions in the observed world and a world where the individual has always belonged to a different demographic group.

EXPERIMENT

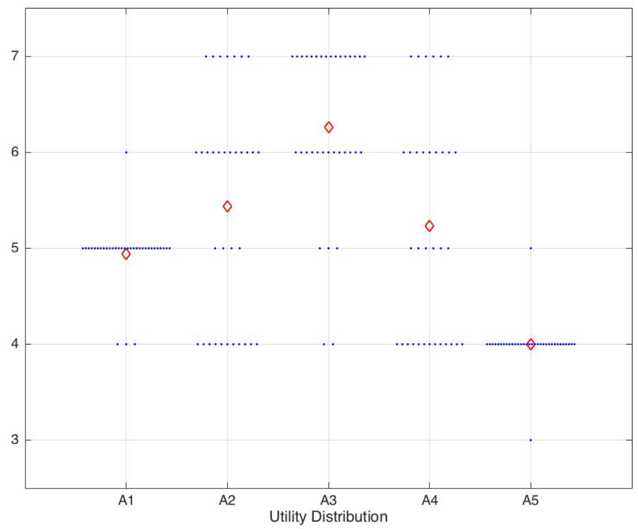
Student's utility = the utility achieved based on the score the student gave to the project the algorithm assigns to her

Student's distance = the total difference between the student's utility and those of all other students, given the projects assigned to everybody by the algorithm

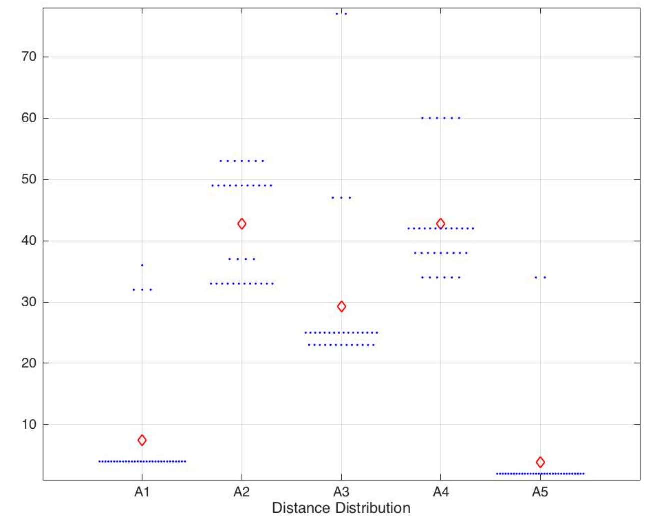
For each algorithm, the table below shows the sum of all student's utilities (total utility) and the sum of students' distances for all students (total distance).

	A1	A2	A3	A4	A5
Total Utility	168	185	213	178	136
Total Distance (10^3 *)	0.2520	1.4540	0.9940	1.4520	0.1320

EXPERIMENT



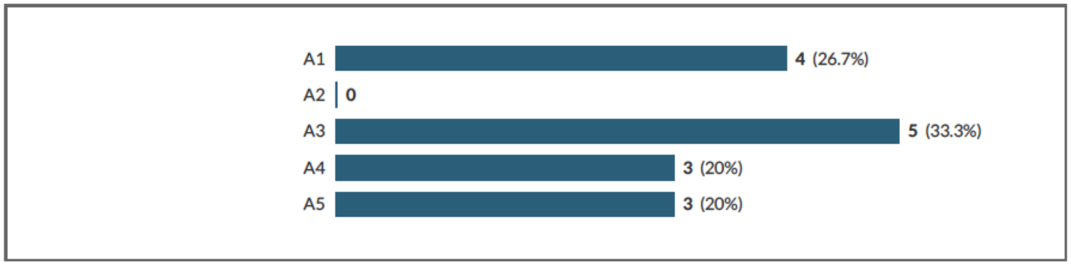
7
6
5
4
3
A1
A2
A3
A4
A5
Utility Distribution



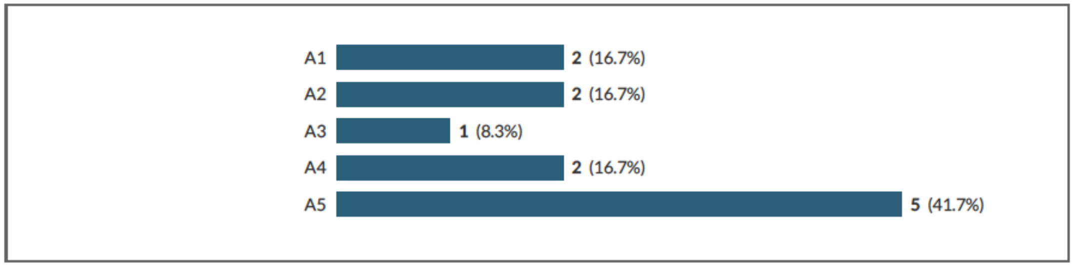
70
60
50
40
30
20
10
A1
A2
A3
A4
A5
Distance Distribution

USER OPINIONS (BEFORE)

Given the allocations computed by each algorithm, which of them would you prefer most? You can list more than one algorithm in each line.



Given the allocations computed by each algorithm, which of them would you prefer least? You can list more than one algorithm in each line.



EXPLANATION

The following is an informal description of how each algorithm works:

Algorithm 1 (A1) minimises the total distance while guaranteeing at least 70% of maximum total utility.

Algorithm 2 (A2) maximises the minimum individual student utility while guaranteeing at least 70% of the maximum total utility.

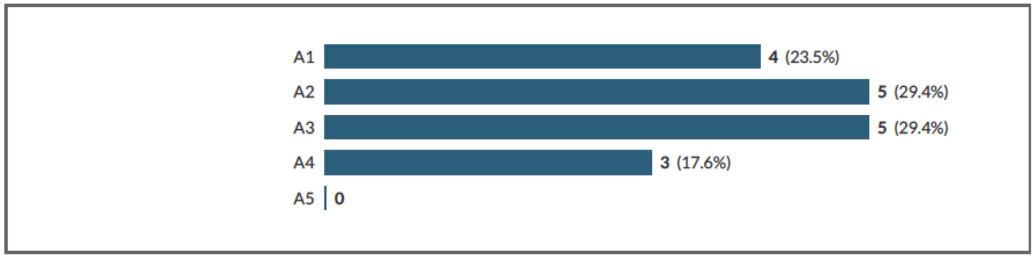
Algorithm 3 (A3) maximises total utility.

Algorithm 4 (A4) maximises the minimum individual student utility.

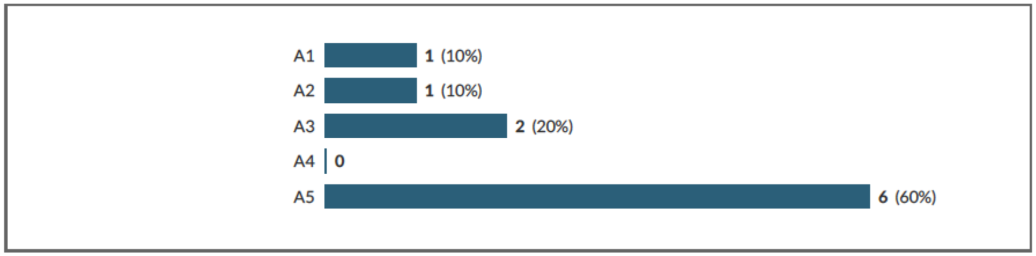
Algorithm 5 (A5): minimises total distance.

USER OPINIONS (AFTER)

Given the explanation of how the algorithms work and the allocations computed by each algorithm, which of them would you prefer most? You can list more than one algorithm in each line.



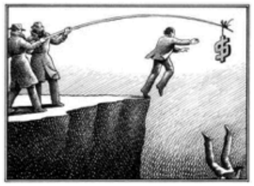
Given the explanation of how the algorithms work and the allocations computed by each algorithm, which would you prefer least? You can list more than one algorithm in each line.



CONCLUSION

- Smart societies should be fair both for **ethical** reasons and to be **sustainable**
- **Algorithmic fairness** itself is a complex and poorly understood notion
- Broader debate is needed to establish solid theoretical framework
- Our (small) contribution
  - Multiagent systems view captures distribution of "wealth", not just statistical properties
  - Might provide models that capture all stakeholders' objectives better
  - Empirical research to understand human notions of fairness
- Is the complexity of solving the general problem the same as that of "optimal political economy" for a globalized society?

WHAT WILL IT BE?

Promise		*Peril*	
Man-machine collaboration		Manipulation	
Personalisation		Surveillance	
Collective intelligence		Humans as cheap labour