

On the Value of Private Information

Jon Kleinberg

Christos H. Papadimitriou

Prabhakar Raghavan

Presentation by *Maurizio Atzori*

atzori@di.unipi.it

What is Privacy?

What is Privacy?

- one of society's most vital concerns

What is Privacy?

- one of society's most vital concerns
- central for e-commerce

What is Privacy?

- one of society's most vital concerns
- central for e-commerce
- arguably the most crucial and far-reaching current challenge and mission of CS

What is Privacy?

- one of society's most vital concerns
- central for e-commerce
- arguably the most crucial and far-reaching current challenge and mission of CS
- least understood scientifically (e.g., is it rational?)

Some thoughts on privacy

Some thoughts on privacy

- also an economic problem

Some thoughts on privacy

- also an economic problem
- surrendering private information is either good or bad for you

Some thoughts on privacy

- also an economic problem
- surrendering private information is either good or bad for you
 - privacy vs. search costs in computer purchasing

Other thoughts on privacy

Other thoughts on privacy

- personal information is intellectual property controlled by others, often bearing negative royalty

Other thoughts on privacy

- personal information is intellectual property controlled by others, often bearing negative royalty
 - such as dinner interruptions and spam

Other thoughts on privacy

- personal information is intellectual property controlled by others, often bearing negative royalty
 - such as dinner interruptions and spam
- selling mailing lists vs. selling aggregate information: false dilemma

Other thoughts on privacy

- personal information is intellectual property controlled by others, often bearing negative royalty
 - such as dinner interruptions and spam
- selling mailing lists vs. selling aggregate information: false dilemma
- Proposal: *Take into account the individual's utility when using personal data for decision-making*

The idea

Take into account the individual's utility when using personal data for decision-making

The idea

Take into account the individual's utility when using personal data for decision-making

- it is possible to precisely quantify that utility, by modeling the value of private information

The idea

Take into account the individual's utility when using personal data for decision-making

- it is possible to precisely quantify that utility, by modeling the value of private information
 - *transaction as coalitional game*

The idea

Take into account the individual's utility when using personal data for decision-making

- it is possible to precisely quantify that utility, by modeling the value of private information
 - *transaction as coalitional game*

Three examples:

- Marketing Survey, Recommendation Systems, Collaborative Filtering

Collaborative Game Theory

We have a set S of n agents and we want to give a value to each agent

Collaborative Game Theory

We have a set S of n agents and we want to give a value to each agent

- For each set $F \subseteq S$ we know the value $v(F)$

Collaborative Game Theory

We have a set S of n agents and we want to give a value to each agent

- For each set $F \subseteq S$ we know the value $v(F)$
- the value $v(S)$ is the sum we want to distribute over the agents

Collaborative Game Theory

We have a set S of n agents and we want to give a value to each agent

- For each set $F \subseteq S$ we know the value $v(F)$
- the value $v(S)$ is the sum we want to distribute over the agents

the **Core** and the **Shapley value**

First idea (notion of “fairness”): the Core

A vector (x_1, x_2, \dots, x_n) with $\sum_i x_i = v(S)$ is in the core if for all $F \subseteq S$ we have $x[F] \geq v(F)$

First idea (notion of “fairness”): the Core

A vector (x_1, x_2, \dots, x_n) with $\sum_i x_i = v(S)$ is in the core if for all $F \subseteq S$ we have $x[F] \geq v(F)$

Problem: *Core is often empty*

Second idea: the Shapley value

$$x_i = \frac{1}{n!} \sum_{\pi} (v[\{j : \pi(j) \leq \pi(i)\}] - v[\{j : \pi(j) < \pi(i)\}])$$

(Meaning: Assume that the agents arrive at random. Pay each one his/her contribution. Average over all possible orders of arrival)

Market Survey

k possible versions of a product

Each customer has exactly one choice

$m = \# \text{majority}$

Market Survey

k possible versions of a product

Each customer has exactly one choice

$m = \# \text{majority}$

Formulation as a coalitional game:

- $n + 1$ agents (agent 0 is the vendor)
- $0 \in F$ then $v(F) = 2 * m$

Market survey problem: the Core

- Suppose largest minority is r

Market survey problem: the Core

- Suppose largest minority is r
- An allocation is in the core as long as losers get 0, vendor gets $\geq 2r$, winners split an amount up to twice their victory margin

Market survey problem: the Core

- Suppose largest minority is r
- An allocation is in the core as long as losers get 0, vendor gets $\geq 2r$, winners split an amount up to twice their victory margin
- (plus another technical condition saying that split must not be too skewed)

Market survey problem: Shapley value

- Suppose margin of victory is at least $\varepsilon > 0\%$ (realistic, close elections never happen in real life)
- Vendor gets $m(1 + \varepsilon)$
- Winners get $1 + \varepsilon$
- Losers get ε (and so, no compensation is necessary)

Recommendation System

- Each participant i knows a set of items B_i
- Each benefits 1 from every new item
- Core: empty, unless the sets are disjoint!
- Shapley value: For each item you know, you are owed an amount equal to $1 / (\text{\#people who know about it})$

(Personal) Comments and Conclusions

- Actually the dissemination is not always detrimental to the individual's interests

(Personal) Comments and Conclusions

- Actually the dissemination is not always detrimental to the individual's interests
 - but the uncontrolled dissemination, it is!

(Personal) Comments and Conclusions

- Actually the dissemination is not always detrimental to the individual's interests
 - but the uncontrolled dissemination, it is!
- The *Shapley value* and *Core* are difficult to find, and (in some contexts) they need some information we do not know

(Personal) Comments and Conclusions

- Actually the dissemination is not always detrimental to the individual's interests
 - but the uncontrolled dissemination, it is!
- The *Shapley value* and *Core* are difficult to find, and (in some contexts) they need some information we do not know
- The “*aggregate data*” approach is (in my point of view) the current solution to avoid the uncontrolled distribution of “very important” private information

References (1/2)

J. KLEINBERG, C. H. PAPADIMITRIOU, P. RAGHAVAN *On the Value of Private Information*, in Proc. of TARK: Theoretical Aspects of Reasoning about Knowledge (2001)

G. CALZOLARI, A. PAVAN, *Optimal design of privacy policies*, manuscript, January 2001

W. HOEFFDING, *Probability inequalities of Sums of Bounded Random Variables*, J.Amer.Stat.Assoc. (1963)

K. C. LAUDON, *Markets and privacy*, Comm. ACM, 36, (1996)

References (2/2)

A. MAS-COLELL, M. D. WINSON, J. R. GREEN *Microeconomic Theory*, Book, Oxford (1995)

M. OSBORNE, A. RUBINSTEIN, *A course in Game Theory*, MIT Press (1994)

P. SAMUELSON, *Privacy as Intellectual Property?*, in *Stanford Legal Review*, 52, Special issue on Privacy (July 2000)

H. VARIAN, *Economic Aspects of Personal Privacy*, Privacy and Self-Regulation in the Information Age, NTIA report