On Dealing with Adversaries Fairly

Andrei Serjantov, Ross Anderson

University of Cambridge
Computer Laboratory

Outline

- The problem (think voting)
  - Why straightforward solutions fail
- Social Choice Theory
  - Arrow & U,D,I,P
- Analysis of Reputation Systems
  - 2 examples
- Some Voting Schemes & Applications
- Manipulability
- Conclusions

A Common Problem

- There are N people wanting to take a decision; each has a set of preferences over the alternatives. e.g.
  - Which is the best? (Elections)
  - Which are bad? (Reputation Systems)
  - What should we do? (Distributed Algorithms)

Deriving a global set of preferences over the alternatives from the individual preferences

How?

- One Person Decides
- Everyone Follows the Party Line
  - Or it’s obvious anyway
- In general, Preference Aggregation
  - (essentially generalized voting)

50 Years of literature on it, do Computer Scientists have to reinvent it all over again???

Democracy (Straightforward Voting) Does Not Always Work

- Just vote over the alternatives & count the votes?
- Suppose we have a cake, and everyone wants as much as possible
- Divide cake equally
- Now vote to take George’s piece and divide between the remaining people

The Cake Division Paradox!

Preference Aggregation Formally

- Have a set X of choices
- Each individual i has preferences defined by the order R_i (a relation which is reflexive, transitive, complete)
- A preference aggregation scheme is a function $f : \{R_i\} \rightarrow R$
- Social Welfare Function (SWF) when the resulting set of preferences R is an order
- Now can state properties of PA schemes

Criteria for Good and Bad

- Suppose we have a preference aggregation scheme, is it good or bad?
- Economists have come up with various criteria over the years … e.g.
  - When one person decides the outcome (dictatorship), this is bad
  - A scheme should “work” for all inputs which it is given (obvious to Computer Scientists, perhaps?)

Properties

- Non-dictatorship
  - No individual dominates choice of x vs y
- Unrestricted Domain
  - The PA always yields an order
- Pareto
  - If everyone says x is better than y, it is!
- Independence of Irrelevant Alternatives
  - Choice between x and y depends on what people say about x vs y, not x vs z and y vs z

Arrow, 1951

- There is no SWF which satisfies all 4 properties

- Followed by many positive results including:
  - if we merely require to be able to choose a set of best alternatives out of the final preferences, a PA scheme (SDF) satisfying the 4 properties exists

Democracy (Straightforward Voting) Does Not Always Work

- Method of Majority Decision does not satisfy Unrestricted Domain
- A: Labour > Conservative > Liberal
- B: Conservative > Liberal > Labour
- C: Liberal > Labour > Conservative
- Hence:
  - Conservative > Liberal (A & B)
  - Liberal > Labour (B & C)
  - Labour > Conservative (A & C)
- Hence R is not transitive!

Properties are Useful

- The properties identified in the literature are useful in thinking about new schemes
- We took a few reputation schemes and tried to determine which properties they have

Reputation Systems

Reputation Systems I
Aberer,Despotovic

- X is the set of all nodes
- $c(x, y)$ is the number of complaints filed by x against y
- The reputation is the number of complaints against x times the number of complaints by x
  - Lower score = higher reputation
- Reputation defined as $T(x) = \sum_i c(x,i) \times \sum_i c(i,x)$
- Reputation remains high (T(x) is 0)
  - Old lady effect (proper, complains all the time)
  - Partisan effect (does bad stuff, never complains)

A fix

- Hence the reputation system on the previous slide is *useless*
- Suppose we add 1 to both sides of the product:

$$T(p) = \sum_i (1 + c(i,x)) \times \sum_i (1 + c(x,i))$$

Does not have an obvious flaw. But how good is it?

Define an Order

$$xR_i y \Leftrightarrow$$

$$\left(1+c(i,x)\right)\times\left(1+c(x,i)\right)\le\left(1+c(i,y)\right)\times\left(1+c(y,i)\right)$$

So $xRy \Leftrightarrow$

$$\sum_i \left(1+c(i,x)\right)\times\sum_i \left(1+c(x,i)\right)\le$$

$$\le \sum_i \left(1+c(i,y)\right)\times\sum_i \left(1+c(y,i)\right)$$

Properties of the New Scheme

- Not Pareto
- Unrestricted Domain
- Independence of Irrelevant Alternatives
- Non-dictatorship

(Clearly better than the old one)

(Involves essentially interpersonal comparisons, but in this case we can cope with that)

Another Reputation Scheme

- Proposed by Dellaroccas (ACM EC 2003)
- Users rate each other on a scale of 1...100
- Ratings are added up, outliers detected and removed
  - Ignores some of the ratings!
  - Involves inter-personal comparisons!
  - Ratings are 1…100 which is unlikely to yield true preferences as users are too confused!
- e.g. Median would have been better!!!

Non-obvious Schemes

- Just to convince you that there are some non-obvious results out there.
- Suppose you have a situation where there is a true global ordering.
- Suppose further that every individual can decide between any x and y whether x<y (in the true ordering) with probability >0.5
- Then there is a scheme (Kemeny-Young) which computes the most probable ordering.
  - (Satisfies the Condorset criterion)

Manipulability

- Gibbard-Satterthwaite Theorem
- Every Voting Scheme is either Dictatorial or Manipulable (by voters)
  - Extended by Gardenfors to orderings
- But how much knowledge do we need to manipulate it?
  - see Nurmi, [20]
- Manipulation might be computationally hard!

Manipulability II

- A recent result shows that voting schemes also suffer from manipulation by candidates entering and leaving the election
  - Well-known in the real world, e.g. recent French elections
- In this framework candidates were also allowed to vote
  - Particularly interesting to Computer Scientists

Conclusions

- Various problems currently considered by Computer Scientists have already been looked at by Economists
- Lots of negative results
- Several Tools directly borrowable
  - Kemeny-Young scheme
- Lessons to be learnt
  - Take care with inter-personal comparisons