The Price of Truth: Minimum payments that make truthful-telling rational

Radu Jurca and Boi Faltings

Artificial Intelligence Laboratory

radu.jurca@epfl.ch
Online Reputation Mechanisms

- become increasingly popular
- several clients interact repeatedly with the same service provider
- feedback about past service => reputation of the provider => future performance
Online Reputation Mechanisms

• become increasingly popular
• several clients interact repeatedly with the same service provider
• feedback about past service => reputation of the provider => future performance
Online Reputation Mechanisms

- become increasingly popular
- several clients interact repeatedly with the same service provider
- feedback about past service => reputation of the provider => future performance
Honest Feedback

• should not be assumed, because reporting is costly

• reporters
  – are altruists (fewer per site)
  – external incentives

• external incentives
  – psychological (extremely satisfied or extremely unhappy clients)
  – material!! (e.g. Amazon, TripAdvisor, eBay)
Reward Feedback Reports

- feedback is paid for (by the Reputation Mechanism)
- payments should be designed such that truth-telling is rational
- Ex-Ante Nash Equilibrium
  - if you don’t know what other agents report, but assume they report honestly, it becomes optimal for you to report the truth
Payment Schemes

- pay the red agent depending on her own report, and the report of another agent (the blue agent)

\[ \tau(s_i, s_j) \]

- possible to find \( \tau(\cdot, \cdot) \) such that: \( \mathbb{E}[\text{Pay}|\text{truth}] > \mathbb{E}[\text{Pay}|\text{lie}] \)
  - because different observations trigger different beliefs for the red agent
Payment Scheme - Example

• you need a plumber!

• Prior expectation:
  – Pr[good service] = 75%
  – Pr[bad service] = 25%

• Posterior beliefs:
  – 😊 when happy
    • Pr[good service | good experience] = 86%
    • Pr[bad service | good experience] = 14%
  – 😞 when unhappy
    • Pr[good service | bad experience] = 43%
    • Pr[bad service | bad experience] = 57%
Example (continued)

- observes 0:
  \[ \Pr[0|0] = 57\%, \quad \Pr[1|0] = 43\% \]

- observes 1:
  \[ \Pr[0|1] = 14\%, \quad \Pr[1|1] = 86\% \]

- observes 1 and reports 1:
  \[ E[\text{Pay}] = 0.86 \times 2 + 0.14 \times 0 = 1.72 \]

- observes 1 and reports 0:
  \[ E[\text{Pay}] = 0.86 \times 1.5 + 0.14 \times 2 = 1.57 \]

- observes 0 and reports 1:
  \[ E[\text{Pay}] = 0.43 \times 2 + 0.57 \times 0 = 0.86 \]

- observes 0 and reports 0:
  \[ E[\text{Pay}] = 0.43 \times 1.5 + 0.57 \times 2 = 1.78 \]
Designing Payment Schemes

• incentive-compatible payments can always be found!

• \( f \ ( \text{posterior beliefs} ) = \text{payment table} \)
  – \( f \) is a scoring rule (nice functions, easy to compute)

• Problem:
  – the margin for truth-telling is small compared to the payments
  – \( \mathbb{E}[\text{Pay}] \sim 1.5; \)
  – \( \mathbb{E}[\text{Pay}|\text{truth}] - \mathbb{E}[\text{Pay}|\text{lie}] \sim 0.15 \)
Designing Payment Schemes (2)

• Assume falsely reporting negative feedback brings you 0.01
  \[ \Rightarrow \text{payments can be as high as 0.10} \]
  \[ \Rightarrow \text{IMPRACTICAL} \]

GOAL:

• decrease the payments as much as possible, keeping desired margins for truth-telling.
Solution 1 – Automated Mechanism Design

• **Intuition:** the values in the payment table are variables in an optimization problem that:
  – minimizes the required budget,
  – makes sure that honesty is rewarded more than lying by at least a given margin.

• linear optimization problem (can be solved efficiently)
Optimal Payments - Performance

- logarithmic scoring rule
- quadratic scoring rule
- spherical scoring rule
- optimal payment scheme

average cost

number of signals (M)
Solution 2 - Several Reference Reports

- use the reports of several agents to compute payments

<table>
<thead>
<tr>
<th></th>
<th>00..0</th>
<th>…</th>
<th>11..1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>…</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>…</td>
<td>1</td>
</tr>
</tbody>
</table>

- the expected cost decreases with the number of reference reports

- however, the complexity increases
Several Reference Reports

The graph shows the average cost as a function of the number of signals (M) for different values of N. The lines represent:

- N = 1 (dotted line with circles)
- N = 2 (dashed line with stars)
- N = 3 (solid line with diamonds)
- N = 4 (dashed-dotted line with squares)
- N = 5 (solid line with triangles)

The x-axis represents the number of signals (M), ranging from 2 to 8, while the y-axis represents the average cost, ranging from 0.5 to 0.95.
Solution 3 - Filtering Feedback Reports

- **Intuition**: pay all feedback reports, but only consider some of them when computing reputation information.
  - efficient filtering mechanisms decrease the incentives for lying (hence the margins required for truth-telling)

- designed automatically with the payment system

| Probability of Filtering Out the Report When \( \tilde{r} \) Reports 0 and the Filtering Reports are \( 00..0 \) |
|---|---|---|
| \( \pi(0,00..0) \) | ... | 
| \( \pi(1,00..0) \) | ... |
Effects of the Filtering Mechanism

M = 5 feedback signals
Convergence

(a) $M = 5$, $\hat{N} = 3$, $\gamma = 0.05$

(b) $M = 5$, $\hat{N} = 3$, $\gamma = 0.15$

(c) $M = 2$, $\hat{N} = 2$, $\gamma = 0.1$

(d) $M = 2$, $\hat{N} = 2$, $\gamma = 0.1$
Reporting Equilibria

• several reporting equilibria exist

• can be eliminated with trusted information

• temporary interventions
  – honest reporting is ESS: can only be changed by a significant group of colluders

• design payments that are resistant to given coalition size
Application – QoS Monitoring

- without QoS monitoring, selfish providers deliver minimum quality

- existing techniques:
  - central monitor (bottleneck)
  - provider (client) side monitoring (not trusted)

- monitoring based on feedback
  - rational incentives to report the truth, (instead of hard security)
Future Work

• reporters have private information
  – more complex optimization problem (worst case analysis)
  – tractable solution (Zohar & Rosenschein, 2006)
  – payments increase

• subjective quality

• providers take strategic decisions or change their type
Conclusions

- it is possible to convince rational agents to report the truth
- use the available computation power to design the minimum payments
- using filtering and payment mechanisms together decreases the payments
- important applications for decentralized markets