Proper Scoring Rules

Boi Faltings

AI Laboratory, EPFL
boi.faltings@epfl.ch
http://liawww.epfl.ch/

2016

Prediction Tasks

Center would like to know/predict the value of phenomenon $P \in \{x_1, .., x_N\}$

Examples:

- Is crop diseased? $x_1 = disease_1, .., x_N = nodisease$
- Who will win the election? $x_1 = cand_1, .., x_N = cand_n$
- How much will this product sell? $x_i = \$i * 1mio$

Only agents can observe phenomenon (or aspects related to it).

Ground truth $g \in \{x_1, .., x_N\}$ will be available later.

Prediction Process

1. Agent has prior belief $Pr(x_1), .., Pr(x_N)$
2. Agent observes phenomenon $o \in \{x_1, .., x_N\}$
3. Agent forms posterior belief $Pr(x_1|o), .., Pr(x_N|o)$
4. Agent reports $A = x_k$ or $A =$ entire posterior belief.

Incentives

Depends on what is being reported.

- value with highest probability: reward iff report matches ground truth.
- posterior distribution: agent should believe that truthfully reporting posterior gives highest payoff.

$\Rightarrow$ use proper scoring rule

Proper scoring rules

Payment function $pay(A, g)$ such that:

$$(\forall Pr') \sum_{x_i inP} Pr(x_i|o) pay(Pr(x_i|o), g) \geq \sum_{x_i inP} Pr(x_i|o) pay(Pr'(x_i|o), g)$$

Examples:

- Quadratic scoring rule:

$$pay(A, g) = 2A(g) - \sum_{x \in P} A(x)^2$$

- Logarithmic scoring rule:

$$pay(A, g) = C + \ln A(g)$$

Prediction Mechanism with Proper Scoring Rules

"What will be the weather next Sunday: Rain, Cloudy or Sunny?"

- Agent prior = historial averages (for example):

$$Pr(P) = \begin{array}{c|c|c} Rain & Cloud & Sun \\ \hline 0.2 & 0.3 & 0.5 \end{array}$$

- Agent studies data $\Rightarrow$ posterior belief:

$$Pr(P|o) = \begin{array}{c|c|c} Rain & Cloud & Sun \\ \hline 0.8 & 0.15 & 0.05 \end{array}$$

- Agent reports a forecast (rain), or the entire posterior belief.
- on Sunday, it rains, and he gets paid:
  - reward C for the correct forecast, or
  - $C + \ln 0.8 = C - 0.22$ using the logarithmic scoring rule.

Why is this truthful?

Expected reward using log scoring rule:

$$E[pay(A,g)] = \sum_{x \in P} Pr(x|o) pay(A,x) = \sum_{x \in P} Pr(x|o) C + \ln(A(x))$$

and the difference between truthful/non-truthful reporting:

$$\begin{aligned} E[(pay(Pr,g)] & -E[pay(A,g)] \\ = & \sum_{x \in P} Pr(x|o)(C + \ln Pr(x)) - (C + \ln A(x)) \\ = & \sum_{x \in P} Pr(x|o) \ln \frac{Pr(x)}{A(x)} \\ = & D_{KL}(Pr||A) \end{aligned}$$

By Gibbs' inequality, $D_{KL}(Pr||A) \geq 0$, so truthful reporting gets the highest possible expected payoff.

Properties of Proper Scoring Rules

Advantages:

- Truthtelling is the dominant strategy.
- Can be used to incentivize effort as well.

Disadvantages:

- Not clear how to aggregate distributions into a joint estimate $\Rightarrow$ prediction markets.
- Ground truth must become known before paying rewards $\Rightarrow$ peer consistency.

Readings

G.W. Brier: Verification of Forecasts Expressed in Terms of Probability, *Monthly Weather Review*, 78, 13, 1950

L.J. Savage: Elicitation of Personal Probabilities and Expectations, *Journal of the American Statistical Association*, 66, 783801, 1971

T. Gneiting, A.E. Raftery: "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102.477, pp. 359-378, 2007