Learning from a Mixture of Information Sources

Nicole Immorlica[1] Brendan Lucier[1] Clayton Thomas[1] Ruqing Xu[2]

[1]Microsoft Research [2]Cornell University

Paper Me

Research Question

Whether a piece of information—such as a product review, a news article, or a medical recommendation—is informative depends not only on its **content** but on **who produces it**. While technologies can summarize large data at low costs, the data **source** is often lost.

How does the value of knowing a signal's source compare with the ability to process more samples?

Model in the Two-signal, No-fake-data Setting*

- $\Theta = \{0, 1\}$ binary *states* endowed with a uniform prior.
- $\Omega = \{0, 1\}$ binary realization space. $\omega \in \Omega$ is a *signal realization*.
- A *signaling scheme* π is a pair (p_{00}, p_{11}), where $p_{\theta\omega} = \pi(\omega \mid \theta)$. $\mathcal{P}$ is the *domain* of feasible signaling schemes.
- In our main result, we focus on the domain of "no-fake-data" signaling schemes $\mathcal{P}_{\triangleleft} = \{(p_{00}, p_{11}) \mid p_{00} + p_{11} \geq 1\}$.

	Signal space Ω: 0	1
State space Θ: 0	p_{00}	$p_{01} = 1 - p_{00}$
1	$p_{10} = 1 - p_{11}$	p_{11}

Table 1. Binary signaling scheme

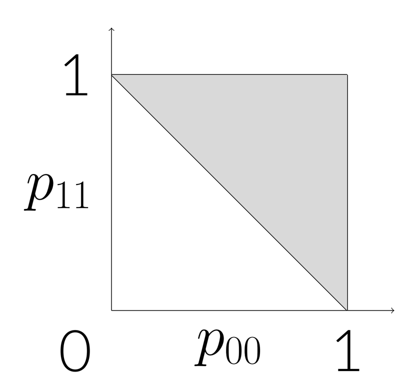

Figure 1. $\mathcal{P}_{\triangleleft}$ domain

Learning: Nature draws a state $\theta \in \Theta$ according to the uniform prior. The decision maker learns about the state from one or more signaling schemes repeatedly drawn from a **distribution** Π over the domain $\mathcal{P}$. Each time, nature draws a signaling scheme $\pi \sim \Pi$, and then a signal realization $\omega \sim \pi(\cdot \mid \theta)$.

Source-aware signal $A(\Pi)$: The decision maker learns the tuple (π, ω), i.e., the signaling scheme and a realization from it.

Source-blind signal $B(\Pi)$: Nature draws θ, π, and ω exactly as before, but the decision maker only learns ω and not π.

Source-aware and Source-blind Learning*

Proposition 1 (Source-blind learners learns the mean signal). For any distribution of signaling schemes $\Pi \in \Delta(\mathcal{P})$, we have that $B(\Pi)$ is equivalent to the "mean signal" $\overline{\pi} = \mathbb{E}[\Pi]$.

Proposition 2 (Source-aware learners are risk-loving in information). For $\Pi_s, \Pi \in \Delta(\mathcal{P})$, suppose Π_s is a mean-preserving spread of Π. Then, $A(\Pi_s)$ Blackwell dominates $A(\Pi)$.

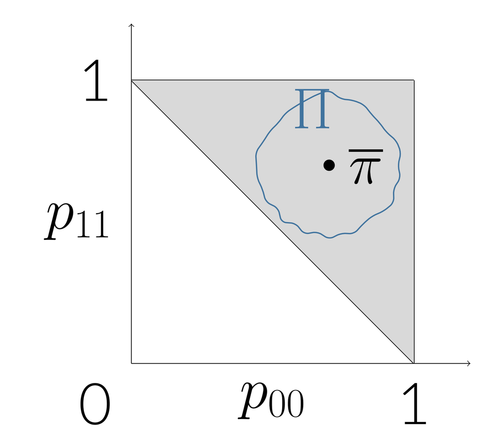

Figure 2. Proposition 1

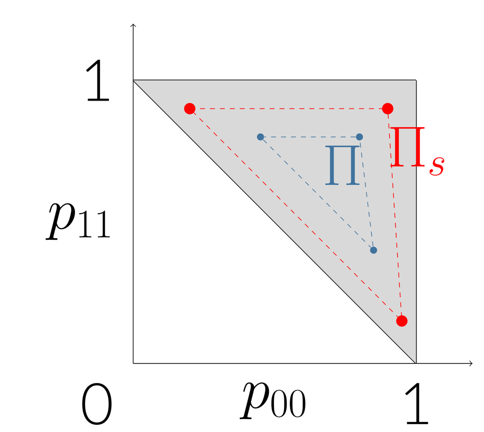

Figure 3. Proposition 2

Over-provisioning Theorem

How much more informative is $A(\Pi)$ than $B(\Pi)$? We put an upper bound on the the dominance ratio $A(\Pi)/B(\Pi)$, as introduced in [2] (see Technical References).

Theorem 1 (Over-provisioning). Let $\Pi \in \Delta(\mathcal{P}_{\triangleleft})$ be any distribution of signaling schemes with the average signal $\mathbb{E}[\Pi] = (x, y)$. If the average signal is ε-away from being completely uninformative, i.e., $x + y \geq 1 + \varepsilon$ for some $\varepsilon > 0$, then $A(\Pi)/B(\Pi)$ is at most $\frac{2\log(1-\varepsilon)}{\log(1-\varepsilon^2)} = O(1/\varepsilon)$.

Interpretation: If the average signal is not too uninformative, then for any decision problem, a source-blind learner with access to a few times, i.e., $O(1/\varepsilon)$, more signals will outperform a source-aware learner.

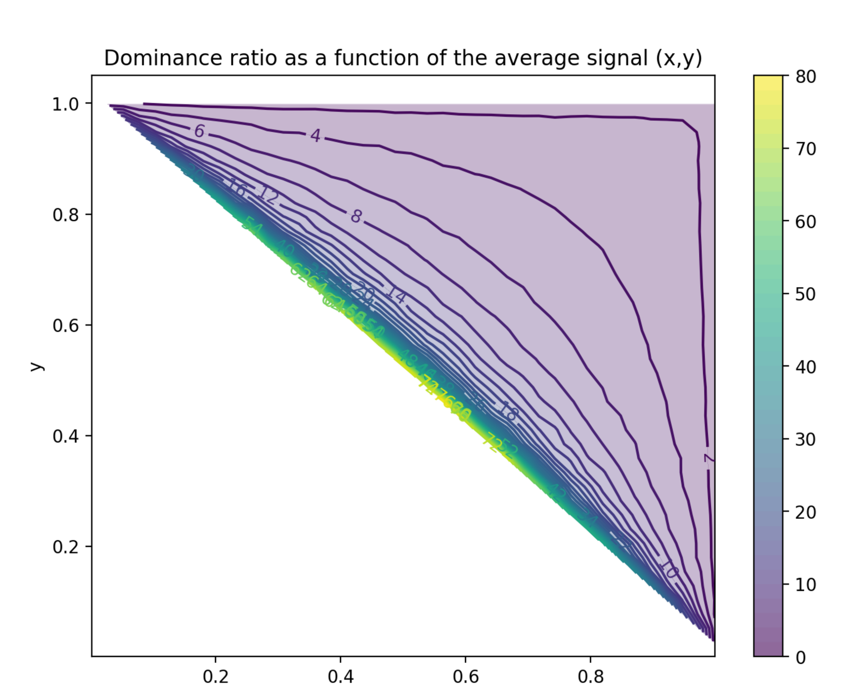

Figure 4. Simulated dominance ratio

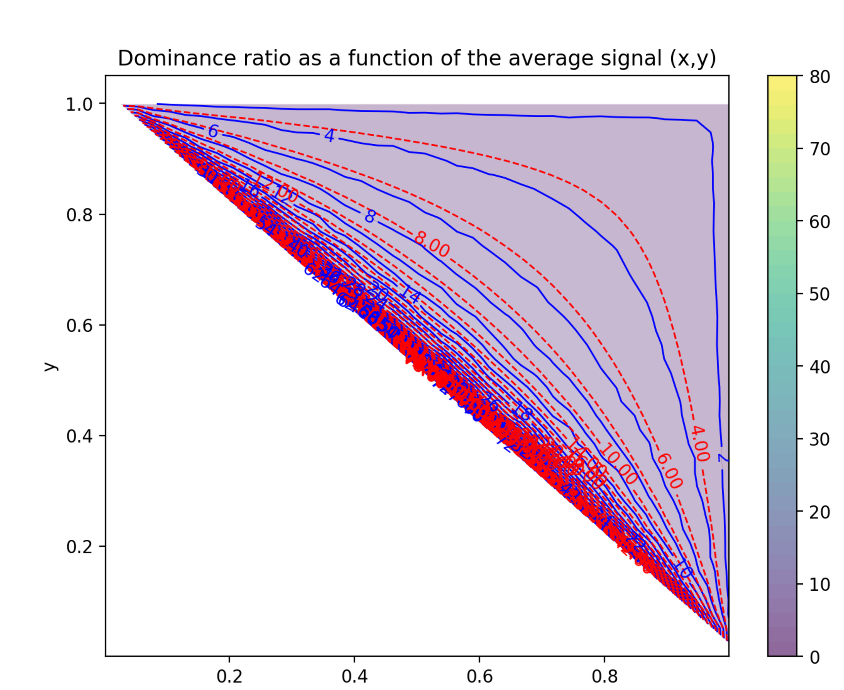

Figure 5. Analytical upper bound

Technical References

Blackwell Experiments

Definition [Distribution over posteriors]. $\tau_\pi \in \Delta(\Delta(\Theta))$ is the distribution over posterior beliefs generated by the signaling scheme π. In binary states, $\tau_\pi \in \Delta([0, 1])$.

Definition [Mean-preserving spread in $\mathbb{R}^m$]. For random variables X_1 and X_2 in $\Delta(\mathbb{R}^m)$, we say X_2 is a *mean-preserving spread* of X_1 if there exists a spread function $s : \mathbb{R}^m \to \Delta(\mathbb{R}^m)$ from X_1 to X_2 such that:

(1) For all t in the support of X_1, we have $\mathbb{E}[s(t)] = t$.

(2) If we draw $z \sim X_1$ and then $y \sim s(z)$, then y is equal in distribution to X_2.

Theorem [1]. Let $P : \Theta \to \Delta(\Omega)$ and $Q : \Theta \to \Delta(\Xi)$ be two signaling schemes with state space Θ and realization spaces Ω and Ξ. The following statements are equivalent:

(i) The posterior distribution τ_P is a mean-preserving spread of τ_Q.

(ii) For every decision problem with state space Θ, any Bayesian decision maker can achieve weakly higher expected utility under P than under Q.

(iii) Q is a garbling of P.

In this case, we say that P **Blackwell dominates** Q, denoted by $P \succeq Q$.

Dominance Ratio

Definition [2]. The **dominance ratio** of two signaling schemes P and Q is defined as:

$$P/Q = \sup\left\{\frac{m}{n} : P^{\otimes n} \succeq Q^{\otimes m}\right\}$$

where $P^{\otimes n}$ means observing n independent realizations from the signaling scheme P. Intuitively, a dominance ratio of r suggests that P will be at least r times as informative as Q in large samples.

References

*See paper for a model of finite state and signal spaces and Proposition 1 and 2 in the general setting.

[1] David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

[2] Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. From blackwell dominance in large samples to rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021.