Information Design with Predictions

KEEGAN HARRIS, Carnegie Mellon University, USA
ZHIWEI STEVEN WU, Carnegie Mellon University, USA
MARIA-FLORINA BALCAN, Carnegie Mellon University, USA

ACM Reference Format:
Keegan Harris, Zhiwei Steven Wu, and Maria-Florina Balcan. 2024. Information Design with Predictions. 1, 1 (April 2024), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Strategic disclosure of information is ubiquitous in domains such as online marketing, sales, and advertising. However in many settings of interest, the optimal disclosure strategy depends on knowledge about the receiver which may be unknown to the individual disclosing information. We initiate the study of information design with predictions.[1] We focus on Bayesian persuasion [2, 3], a canonical branch of information design. In Bayesian persuasion, a *sender* reveals information about a payoff-relevant state to a *receiver*, who takes a payoff-relevant action. We study a persuasion setting with two states and two actions in which the sender does not know the receiver's prior, but receives a *prediction* about its value. We introduce the notions of *consistent* and *robust* sender strategies for information disclosure, and we provide a sender strategy (i.e., a *messaging policy*) which optimally trades off between the two.

2 SETTING

We consider a Bayesian persuasion game between two players: a sender and receiver. Their interaction proceeds as follows:

(1) The sender and receiver share common prior $p := \mathbb{P}(\omega = 1)$ over possible states $\Omega = \{0, 1\}$.
(2) $\omega \sim p$. Nature sends signal $s \sim \mathcal{D}(\omega)$ to the receiver, where $\mathcal{D}(\cdot)$ has support $\mathcal{S}$. The receiver forms new prior $\pi_s := \mathbb{P}(\omega = 1|s) = \frac{p\mathbb{P}(s|\omega=1)}{\mathbb{P}(s)}$.[2]
(3) The sender receives a prediction of the new prior $\widehat{\pi}_s \in [0, 1]$ and commits to messaging policy $\sigma : \mathbb{R} \times \Omega \to \mathcal{M}$, where $\mathcal{M}$ is the sender's message space.
(4) The sender observes state ω, sends message $m \sim \sigma(\omega; \widehat{\pi}_s)$ to the receiver.
(5) The receiver takes action $a = \arg\max_{a\in\{0,1\}} \mathbb{E}_{\pi_s}[u_R(\omega, a)|m]$, where $u_R(\omega, a) = \mathbb{1}\{\omega = a\}$. The sender receives utility $u_S(\omega, a) = a$.

The following definitions will be useful in the sequel.

[1]See Bergemann and Morris [1] for an overview of the literature on information design.
[2]When it is clear from the context, we will suppress the dependence on s and write π_s as π.

Authors' Contact Information: Keegan Harris, keeganh@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Zhiwei Steven Wu, zhiweiw@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Maria-Florina Balcan, ninamf@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

ACM XXXX-XXXX/2024/4-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Definition 2.1 (Ex-Post Expected Utility). $U_\pi(\sigma(\cdot;\widehat{\pi}))$ is the sender's expected utility under prior π and prediction $\widehat{\pi}$, i.e., $U_\pi(\sigma(\cdot;\widehat{\pi})) := \mathbb{E}_{\omega\sim\pi}\mathbb{E}_{m\sim\sigma(\omega;\widehat{\pi})}[u_S(\omega, a(m,s))]$, where $a(m,s) := \arg\max_{a\in\mathcal{A}} \mathbb{E}_{\omega\sim\pi}[u_R(\omega,a)|m]$.

The following characterization of the sender's optimal messaging policy is well-known in the literature on Bayesian persuasion. Observe that the form of the sender's optimal policy depends on the (unknown) prior π.

Proposition 2.2 (Optimal Messaging Policy). *The sender's optimal messaging policy takes the form* $\sigma^*_\pi := \{\sigma^*_\pi(m=1|\omega=1)=1,\ \sigma^*_\pi(m=1|\omega=0)=\frac{\pi}{1-\pi}\}$. *The ex-post sender utility is* $U_\pi(\sigma^*_\pi) = 2\pi$.

Some of our results will leverage the following *truth-telling* messaging policy, which simply reveals the true state to the receiver.

Definition 2.3 (Truth-Telling Messaging Policy). The truth-telling policy always reveals the state and is denoted by σ_T, i.e., $\sigma_T := \{\sigma_T(m=1|\omega=1)=1, \sigma_T(m=0|\omega=0)=1\}$.

Finally, we define the following notions of consistency and robustness, which our messaging policy will trade off between.

Definition 2.4 (Ex-Post γ-consistency). Messaging policy σ is ex-post γ-consistent if for any π, $\frac{U_\pi(\sigma(\cdot;\pi))}{U_\pi(\sigma^*_\pi)} \geq \gamma$.

Definition 2.5 (Ex-Post β-robustness). Messaging policy σ is ex-post β-robust if for any π, $\widehat{\pi}$, $\frac{U_\pi(\sigma(\cdot;\widehat{\pi}))}{U_\pi(\sigma^*_\pi)} \geq \beta$. Messaging policy σ is ex-post β-robust *with respect to truth-telling* if $\frac{U_\pi(\sigma(\cdot;\widehat{\pi}))}{U_\pi(\sigma_T)} \geq \beta$.

3 OVERVIEW OF RESULTS

Our first result shows that the notion of robustness with respect to the optimal policy is too strong, in the sense that no messaging policy can obtain better robustness than truth-telling, for any level of consistency.

Theorem 3.1 (Impossibility of β-robustness). *When $S > 1$, no messaging policy is better than* $\max\left\{\frac{1}{2},\pi\right\}$*-robust. Moreover, the truth-telling messaging policy attains* $\max\left\{\frac{1}{2},\pi\right\}$*-robustness.*

As a result of Theorem 3.1, we will only consider robustness with respect to truth-telling in what follows. The following lemma identifies several properties about the optimal trade-off between consistency and robustness.

Lemma 3.2. *Any policy which is $(1-\alpha)$-robust can be no better than $(\alpha + (1-\alpha)\max\left\{\frac{1}{2},\pi\right\})$-consistent. Moreover, any policy which is more than $(\alpha+\frac{1-\alpha}{2})$-consistent can be no more than $(1-\alpha)$-robust.*

Our main result is a characterization of a messaging policy which exhibits the optimal trade-off between consistency and robustness. The idea is simple: the optimal trade-off is given by taking a convex combination between the optimal messaging policy when $\widehat{\pi}=\pi$ (Proposition 2.2) and the truth-telling policy (Definition 2.3).

Theorem 3.3. *The following messaging policy exhibits the optimal trade-off between consistency and robustness, i.e. it is at least $\alpha + (1-\alpha)\max\{\frac{1}{2},\pi\}$-consistent and $(1-\alpha)$-robust with respect to truth-telling. Set:* $\sigma(m=\widehat{\pi}|\omega=1)=\alpha, \quad \sigma(m=1|\omega=1)=1-\alpha,$

$$\sigma(m=\widehat{\pi}|\omega=0) = \alpha\min\left\{\frac{\widehat{\pi}}{1-\widehat{\pi}},1\right\}, \quad \sigma(m=0|\omega=0) = 1-\alpha+\alpha\cdot\left(1-\min\left\{\frac{\widehat{\pi}}{1-\widehat{\pi}},1\right\}\right)$$

Observe that in order to achieve this guarantee, the sender does not need to know any information about Nature's signaling policy $\mathcal{D}$.

REFERENCES

[1] Dirk Bergemann and Stephen Morris. 2019. Information design: A unified perspective. *Journal of Economic Literature* 57, 1 (2019), 44–95.
[2] Emir Kamenica. 2019. Bayesian persuasion and information design. *Annual Review of Economics* 11 (2019), 249–272.
[3] Emir Kamenica and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review* 101, 6 (2011), 2590–2615.