Intro to Informational Substitutes

YILING CHEN
Harvard University
and
BO WAGGONER
University of Pennsylvania

We propose definitions for when signals are “substitutes” and “complements”. These give a characterization of equilibria of prediction markets and are relevant to the complexity of information acquisition under constraints.

Categories and Subject Descriptors: F.2.0 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity

General Terms: Economics, Algorithms, Theory

Additional Key Words and Phrases: Prediction markets, Value of information, Information acquisition, Signals, Substitutes, Complements.

1. DEFINING INFORMATIONAL SUBSTITUTES AND COMPLEMENTS

The goal of *Informational Substitutes* [Chen and Waggoner 2016] is to raise the question: When should we consider a set of pieces of information to be “substitutes” or “complements”, and how can such definitions be useful?

The key intuition for “substitutes” that we hope to leverage here is *diminishing marginal value*: A_1 and A_2 are “substitutes” if having either one makes the other less desirable.[1] But while it is understood how to model substitutable items, new challenges are raised by information. How do we model “information” at all? Where does the “value” of information come from? What is a “marginal” unit of information, e.g., is information divisible?

This note will briefly describe a model for value of information, how we use this to define substitutes and complements according to the above intuition, and how these definitions may be applied in prediction markets and algorithms.

Information and its value. In this work, we model information probabilistically, as in Bayesian games. There is a known prior distribution on an event[2] E of interest and on “signals” (events, possibly observable and correlated with E) that might be available. An agent observing one or more signals makes a Bayesian update to a posterior distribution about other signals and E.

Given this framework, what is the “value” or “utility” for a piece of information?

[1]For example, in the case of items, “bread” and “pasta” are considered substitutes for an agent if the marginal value of bread, $v(\text{bread}) - v(\text{nothing})$, becomes smaller if she already has pasta, $v(\text{bread, pasta}) - v(\text{pasta})$. This intuition is captured by *submodular* valuation functions v.
[2]For this note, we think of events as random variables on a finite set of outcomes.

Authors’ addresses: yiling@seas.harvard.edu, bwag@seas.upenn.edu

We follow [Howard 1966]: Consider a single agent facing a *decision problem* of selecting a decision $d \in \mathcal{D}$ to maximize expected utility, which depends on E:

$$\max_{d \in \mathcal{D}} \; \mathbb{E}\, u(d, E)$$

If this agent first observes a signal A, then she may be able to make better decisions. $\mathcal{V}(A)$ denotes the expected utility when observing the signal A prior to acting:

$$\mathcal{V}(A) \;:=\; \mathop{\mathbb{E}}_{a \sim A} \; \max_{d \in \mathcal{D}} \; \mathbb{E}\left[u(d, E) \mid A = a\right].$$

This reads: "the expectation, over observations of A, of the utility for acting optimally given that observation." An example is pictured in Figure 1.

We propose three versions of the definitions, "weak", "moderate", and "strong" substitutes. These respectively correspond to marginal information being captured by an entire additional signal; any deterministic function of a signal; or any randomized function of a signal. The appropriate definition to use depends on the assumptions of the situation, such as strategies available to a strategic agent.

Defining weak substitutes.[3] We define a set of signals $\{A_1, \ldots, A_n\}$ to be *weak substitutes* for a particular decision problem if $\mathcal{V}$ is submodular: for any subsets of signals $A' \subseteq A$ and any other signal A_i,

$$\mathcal{V}(A \cup A_i) - \mathcal{V}(A) \leq \mathcal{V}(A' \cup A_i) - \mathcal{V}(A').$$

Each side is the marginal value of A_i to, respectively, A and A', so the inequality says "the more you know, the less marginally valuable is A_i".

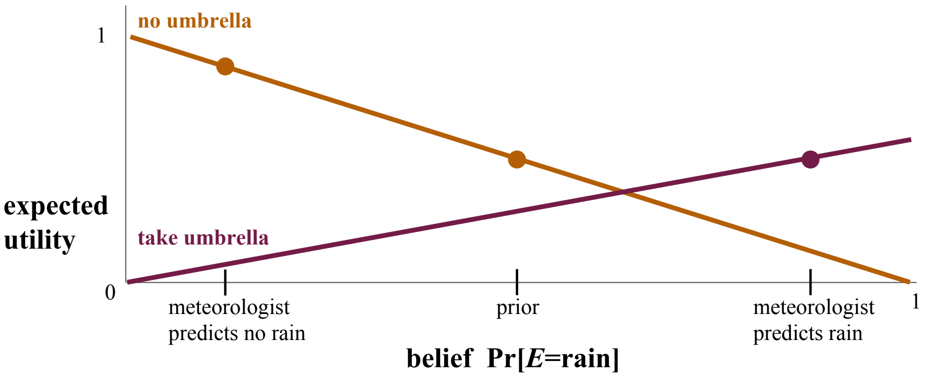
We emphasize that substitutable relationships among signals arise from the interplay of two factors: (1) the information structure of the signals themselves, i.e. how are they correlated with each other and with the event; (2) the utility function or decision problem the agent faces. Except in very rare cases (see the full paper), we cannot conclude substitutability from the information structure alone; signals that are substitutes to one agent may not be substitutes to another.

Defining moderate and strong substitutes. In a strategic setting, the above definitions are not sufficient because, for instance, an agent does not have to choose between either completely revealing a signal A_i or completely hiding it. She can choose to *partially reveal* A_i. For instance, if A_i is a barometer reading, she could choose to e.g. round it to the nearest integer before announcing it. This would give some *less informative* signal $A'_i = f(A_i)$, where f in this example is the rounding function. We use a partial ordering to capture this informativeness and write $A'_i \preceq A_i$.[4] The definition of *moderate substitutes* is exactly the same as for weak, except that the inequality must hold not just for A_i but also hold for all $A'_i \preceq A_i$.

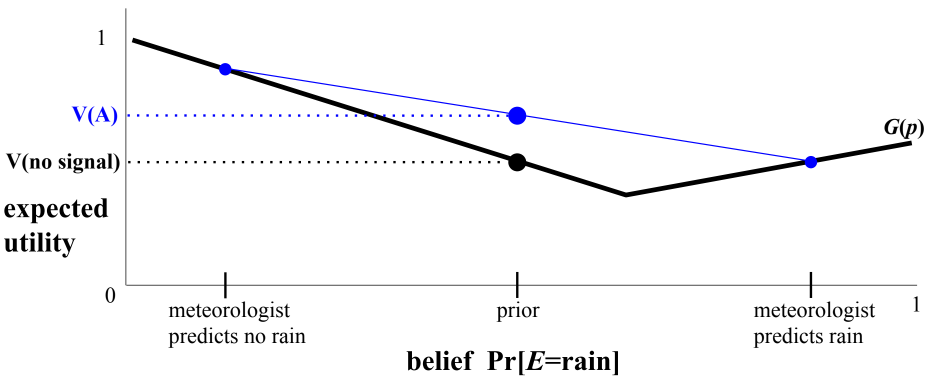
[3]Complements are defined completely analogously using *increasing* marginal value and supermodularity, so we focus on the substitutes case here.

[4]The set of all such signals form a lattice with this partial ordering if using Aumann's partition model [Aumann 1976]. The partial ordering is almost identical to Blackwell's ordering [Blackwell 1953], but here we only care about how much a signal tells the agent about E.

Fig. 1: **Visualizing Value of Information.** An agent, "Elphaba", must decide either to take her umbrella or not. Nature's event $E \in \{\text{rain}, \text{no rain}\}$ uniformly. Let $u(\text{umbrella}, \text{rain}) = 0.6$, $u(\text{no umbrella}, \text{no rain}) = 1$, and otherwise 0. An example signal A is a meteorologist prediction of either rain or no rain, which we suppose is accurate with probability 0.8.



(a) **Expected utility for each action.** For each action, Elphaba faces some (affine!) expected utility as a function of her belief about rain. She will always take the optimal action after observing any available information. E.g. if she observes the meteorologist predicting rain, then she updates to belief $\Pr[E = \text{rain}] = 0.8$ and takes the umbrella; her expected utility given the prediction is $0.8u(\text{umbrella}, \text{rain}) + 0.2u(\text{umbrella}, \text{no rain}) = 0.48$ (purple dot, right). If she observes no signal, her belief remains at 0.5 and she chooses not to take the umbrella, obtaining expected utility $0.5u(\text{no umbrella}, \text{rain}) + 0.5u(\text{no umbrella}, \text{no rain}) = 0.5$ (orange dot, center).



(b) **Value of information.** $G(p)$ traces out the expected utility of the optimal action for each possible belief, and is in a sense sufficient for understanding the decision problem. It is a pointwise maximum of affine functions, hence always convex. $\mathcal{V}(A)$ is Elphaba's expected utility for observing the meteorologist's prediction A, then acting optimally. (It is the average of the cases where he predicts rain or no rain.) The *marginal value* of A over nothing is $\mathcal{V}(A) - \mathcal{V}(\text{no signal})$. This is the amount by which Elphaba expects her utility to increase, on average, if she turns on the weather channel before deciding.

There is one restriction: moderate substitutes assumed that the function f was *deterministic*. However, the agent could even release some randomized function or *garbling*. This gives a much more fine-grained lattice of signals and partial ordering. Signals are *strong substitutes* if the inequality holds for all $A_i' \preceq A_i$ according to this ordering, or in other words, for all $A_i' = f(A_i)$ for any randomized function f.

Connections to (generalized) entropy. One very special decision problem is *predicting against the* log *scoring rule*, where the "decision" is to choose a prediction $q \in \Delta_E$ (a distribution on E) and the score is $u(q,e) = \log q(e)$. It is *proper*, so the optimal action is to report one's posterior and the expected score is $G(q) = \sum_e q(e) \log q(e)$, the negative Shannon entropy of q. In this case, marginal value is equal to mutual information, i.e. $\mathcal{V}(A) = H(E) - H(E|A) = I(E;A)$, and also equals the expected KL-divergence between prior and posterior beliefs. It also follows from prior work that independent signals are complements for the log scoring rule (a fact that generalizes well to other decision problems), while signals conditionally independent on E are substitutes (a fact that does not).

Each decision problem has a corresponding convex $G(q)$, the "expected utility for acting optimally with belief q" (e.g. Figure 1). We can then view $-G(p)$ as a "generalized entropy" (it satisfies some nice entropy axioms), take Bregman divergence of G in place of KL, and the above discussion generalizes.

2. APPLICATIONS

Prediction markets. The goal of a prediction market, given a group of agents with private signals, is to aggregate these into a prediction about an event E conditioned on all of the signals. In "scoring rule" prediction markets, agents take turns making predictions $p^{(1)}, p^{(2)}, \dots$ in a pre-specified order, each observing previous predictions. After the market closes, the event $E = e$ is observed, and an agent who made prediction $p^{(t)}$ is paid by the "improvement in score" that this prediction made over the previous one, according to a proper scoring rule S: namely, $S(p^{(t)}, e) - S(p^{(t-1)}, e)$.

The question we address is under what conditions information is aggregated "as quickly as possible", with agents fully and truthfully reporting their beliefs at the earliest opportunity in equilibrium, as opposed to withholding information. The main result is that the equilibria of the market match this ideal if and only if signals are substitutes. Meanwhile, the "worst-possible" equilibria with maximally-delayed aggregation are characterized by cases where signals are complements.

These results hold for "strong" versions of substitutes and complements, which is appropriate because in general agents may use randomized strategies. (Recall that the definitions depend on a particular utility function of a decision problem; in this case, it is the scoring rule S chosen by the market designer. So these results have some implications for market design.)

Algorithms for information acquisition. In analogy with literature on submodular maximization problems, consider for instance: *There is a set of signals $A_1, \dots, A_n$ available with prices $\pi_1, \dots, \pi_n$. A decisionmaker with a utility function $u(d,e)$ and a budget constraint B must choose a subset of signals to purchase, after which she will observe them and take the optimal action. Which subset should she* *choose?* One can also imagine many other constraint systems than the budget (or "knapsack") constraint above. We call this class of problems SignalSelection.

Interesting questions arise involving how to model the input, but roughly, we show that when the input is accessible efficiently via oracles, SignalSelection reduces both to and from *set function maximization*: choosing a subset S of $\{1, \ldots, n\}$ to maximize some set function $f(S)$ under (to continue the example above) a budget constraint B, where including index i costs π_i.

If and only if the original signals were weak substitutes, f is a submodular set function; if and only if weak complements, a supermodular function.[5] This implies that SignalSelection has polynomial-time approximation algorithms if signals are "weak" substitutes, but does not for complements or in general.[6]

3. KEY PRIOR WORK

In that the paper proposes definitions for informational substitutes and complements, it is a follow-up to [Börgers et al. 2013] and extends the main ideas of that paper. Where we study the game-theoretic equilibria of prediction markets where agents participate multiple times, it is a follow-up to [Chen et al. 2010] and [Gao et al. 2013], which address the key special case of the log scoring rule. Where we study the SignalSelection problem of actively acquiring information under constraints, it follows or extracts ideas from the submodular optimization and information acquisition literatures, see e.g. [Krause and Guestrin 2005; Krause and Golovin 2012].

REFERENCES

Aumann, R. J. 1976. Agreeing to disagree. *Annals of Statistics 4,* 6, 1236–1239.

Blackwell, D. 1953. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics 24,* 2, 265–272.

Börgers, T., Hernando-Veciana, A., and Krähmer, D. 2013. When are signals complements or substitutes? *Journal of Economic Theory 148,* 1, 165–195.

Chen, Y., Dimitrov, S., Sami, R., Reeves, D. M., Pennock, D. M., Hanson, R. D., Fortnow, L., and Gonen, R. 2010. Gaming prediction markets: equilibrium strategies with a market maker. *Algorithmica 58,* 4, 930–969.

Chen, Y. and Waggoner, B. 2016. Informational substitutes. In *56th Annual IEEE Symposium on Foundations of Computer Science*. FOCS '16.

Gao, X. A., Zhang, J., and Chen, Y. 2013. What you jointly know determines how you act: strategic interactions in prediction markets. In *Proceedings of the 14th ACM Conference on Electronic Commerce*. EC '13. ACM, 489–506.

Howard, R. A. 1966. Information value theory. *IEEE Transactions on Systems Science and Cybernetics 2,* 1, 22–26.

Krause, A. and Golovin, D. 2012. Submodular function maximization. *Tractability: Practical Approaches to Hard Problems 3*, 19.

Krause, A. and Guestrin, C. 2005. Near-optimal nonmyopic value of information in graphical models. In *21st Conference on Uncertainty in Artificial Intelligence*. UAI '05.

[5]The "weak" definitions are appropriate because the problem statement requires selecting "all or none" of each signal, not partial information about any signal.

[6]We believe this will be unsurprising and not particularly new to the submodular maximization and information acquisition community. The contribution, if any, lies more in identifying a general formulation of the problem setting and connecting it to the substitutes condition.